G9a, a putative histone methyl-transferase in *Drosophila* interacts with Tungus, a protein associated with α-Actinin

Mona Bjørkmo

Thesis for the
Degree of *Master of Science*

Programme for Molecular
Genetics, Department of
Molecular Biosciences,
University of Oslo, August 2005

Acknowledgements

The work presented in this thesis was carried out at the Department of Molecular Biosciences, University of Oslo. Supervision has been provided by Professor Andrew Lambersson (formal supervisor) and PhD-student Marianne Stabell. Co-supervisor has been Professor Reidunn B. Aalen.

First I would like to thank my supervisor Professor Andrew Lambersson for the support provided during this project, and for assisting me throughout the writing process. Special thanks are in order for allowing me to participate at the Drosophila Genetics and Genomics course in CSHL.

I am especially grateful for Marianne Stabell's excelent introduction into the world of *Drosophila* and molecular methods.

Professor Reidunn B. Aalen and her group at "genetikken" have provided invaluable expertise in the field of epigenetics, in addition to being an entertaining bunch to spend the evenings with.

I wish to thank Roy Falleth and Solveig Hauge Engebretsen for making solutions and media, and Ann Mari Voie for providing an outstanding injection service.

To all my friends and family, thanks for being understanding and supportive.

Last, but not least a big thanks goes to Erling, for making it so easy for me to continue my studies. You have supported and encouraged me in the best ways possible.

Oslo, August 2005

Mona Bjørkmo

Abstract

Histone lysine methylation is considered to be a relatively stable modification associated with important functions in epigenetic gene control and for organizing chromatin domains. Genes encoding mammalian homologues of the *Drosophila* suppressor of PEV *Su(var)3-9* were the first shown to encode proteins with histone lysine methyl-transferase (HKMT) activity. A hallmark signature of this class of proteins is the evolutionary conserved SET-domain found in numerous chromatin regulators, and was named for its occurrence in genes encoding three such regulators in *Drosophila*, namely *Su(var)*3-9, *E(z)* and *trithorax*.

Here we describe the characterization of a putative SET-domain gene in *Drosophila melanogaster*, *G9a*. The gene encodes a protein of 1637 amino acids with similar domain architecture as the mammalian homologue of same name. Whole mount *in situ* hybridization shows that the gene is maternal and immunostaining shows nuclear localization of DmG9a. A yeast two-hybrid screening revealed that DmG9a interacts with Tungus, a LIM-domain protein associated with α-Actinin. Further analysis is needed to investigate the functional implications of this putative interaction.

Table of contents

Abbreviations

aa	**Amino acid**
AD	**Activation domain**
ANK	**Ankyrin**
BD	**Binding domain**
EMS	**Ethyl methane sulphonate**
GST	**Glutathione S-transferase**
HKMT	**Histone lysine methyl-transferase**
HMT	**Histone methyl transferase**
HP1	**Heterochromatin protein 1**
HS	**Hybridization solution**
IPTG	**Isopropyl ß-D-thiogalactoside**
LIM	***Linl-1, Isl-1, Mec-3***
MZT	**Maternal to zygotic transition**
PcG	***Polycomb* Group**
PDZ	***PSD-95, Dlg, ZO-1***
PEV	**Position-effect variegation**
PRMT	**Protein arginine methyl-transferase**
QDO	**Quadruple dropout medium**
RT-PCR	**Reverse transcriptase polymerase chain reaction**
SB	**Staining buffer**
SET	***Su(var)3-9, E(z), trithorax***
SDS-PAGE	**Sodium dodecyl sulphate polyacrylamide gel electrophoresis**
TDO	**Triple dropout medium**
trxG	***trithorax* Group**
ZASP	**Z-band alternatively splized PDZ-motif protein**
Y2H	**Yeast two-hybrid**
ZM	**ZASP-like motif**

1. Introduction

1.1 Chromatin and SET-domain proteins

1.1.1 Packaging of DNA

Eukaryotic cells contain considerable amounts of DNA that has to fit into the nucleus. To accomplish this, nucleotides are tightly folded into a complex called chromatin with the help of specialized proteins called histones that provide increasingly higher levels of organization (Fig. 1.1).

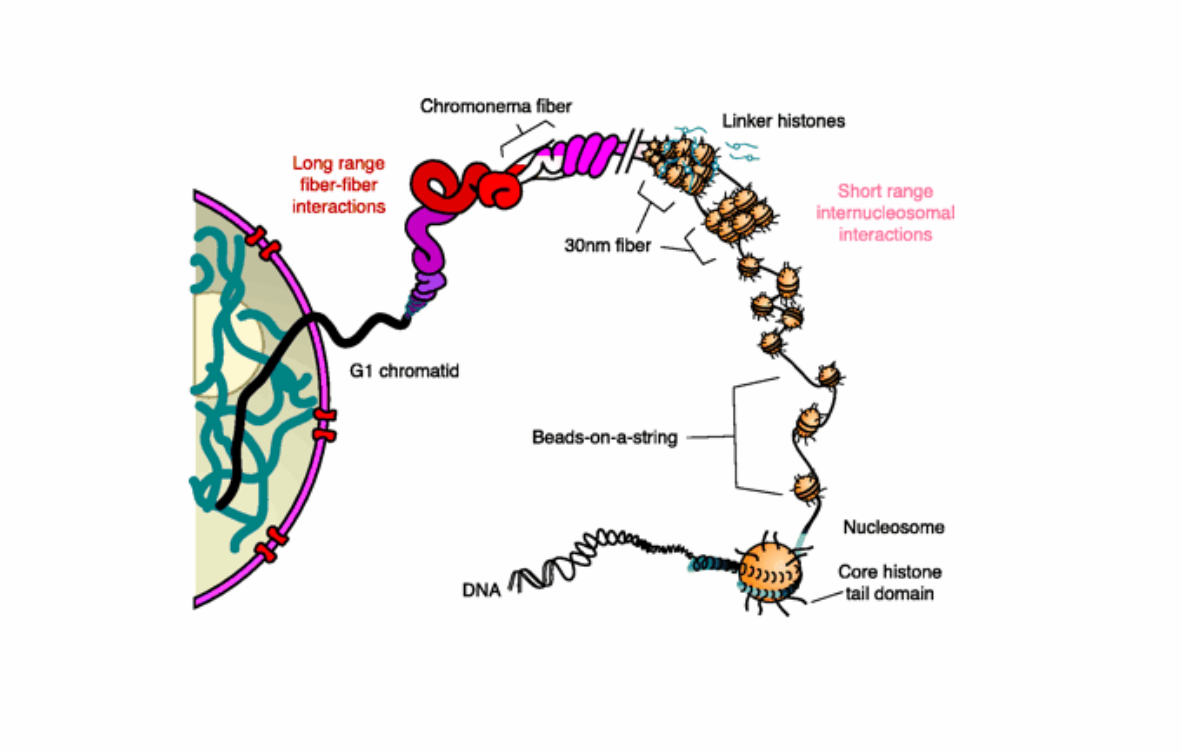

Figure 1.1: Multiple levels of chromatin folding. The basic repeated unit of chromatin is the nucleosome. The nucleosomes are arranged in a "beads on a string" fashion. These are again packaged into a structure called 30nm fiber. Chromatin structure beyond this stage remains uncharacterized. Adapted from Horn and Peterson (2002).

The basic repeated unit of chromatin is the nucleosome. It is composed of a helix of 147 bp DNA wrapped 1.7 times around an octamer of core histones (Luger et al. 1997). Each core nucleosome is composed of two dimers of H2A and H2B and one tetramer of H3-H4. Linker histone, H1, associate with 10-60 bp of DNA between single nucleosomes establishing an increased level of organization, the 30 nm fiber. Conformation beyond this structure is less characterized, and the grade of compaction varies. It is either in a highly compact state called heterochromatin, or in a less tightly packed state called euchromatin. The more closed state of heterochromatin contains few actively expressed genes, whereas euchromatin is associated with active transcription. Chromatin is a highly flexible structure which undergoes dynamic changes during many genetic processes. They include necessary structural reorganizations that occur during DNA replication and cell cycle progression, spatially and temporally coordinated gene expression, as well as DNA repair and recombination events.

Heterochromatin is further differentiated into a constitutive and a facultative state. Constitutive heterochromatin is permanently condensed and is predominantly found at centromeric and telomeric chromosomal regions, as well as in regions containing transposable elements. In contrast, facultative heterochromatin represents transiently condensed and silenced euchromatin. Best known examples are the inactivated X chromosome in female mammals, and position-effect variegation (PEV). PEV as a phenomenon was first described in *Drosophila melanogaster* (Muller 1930) where euchromatic genes in the vicinity of heterochromatic sequences can be become transcriptionally silent due to chromosomal rearrangement or transposition. The extent of silencing varies from one clonal population to another and thus gives rice to the variegated phenotypes. The finding of PEV did not only support the link between heterochromatin and transcriptional silencing, but also suggested that there exist changes that influence phenotype without altering DNA sequence. The study of stable alterations in gene expression potential that not are mediated at the DNA sequence level is called *epigenetics*, and includes post-synthetic modifications on

either DNA itself or of proteins that intimately associate with DNA as the key mediators.

1.1.2 Histone code hypothesis

The N-terminal histone tails are extending outwards from the core of the nucleosomes (Fig. 1.2), and are highly accessible for covalent post-translational modifications such as acetylation, phosphorylation, methylation, ADP-ribosylation and ubiquitination (Ausio et al. 2001). The first three modifications have been studied extensively in the resent years; however, further studies are needed in order to elucidate the complete function of all these modifications (Margueron et al. 2005). Enzymes transferring these characterized modifications are highly specific for particular amino acid positions. They may contribute to an alteration in the interaction of the histone tails with DNA or with chromatin-associated proteins that may be required for different downstream cellular processes. Several years of work on histone modifications and regulation of gene expression lead to the "histone code" hypothesis (Strahl and Allis 2000; Turner 2002). This model postulates that the modifications in different variations specify a language that dictates the regulatory features of a gene, and is read by a set of proteins that translate the code into a particular chromatin state, either active or repressed. It is, however, important to realize that the pattern of histone marks can be differentially interpreted by cellular factors, depending on the gene being studied and the cellular context (Peterson and Laniel 2004).

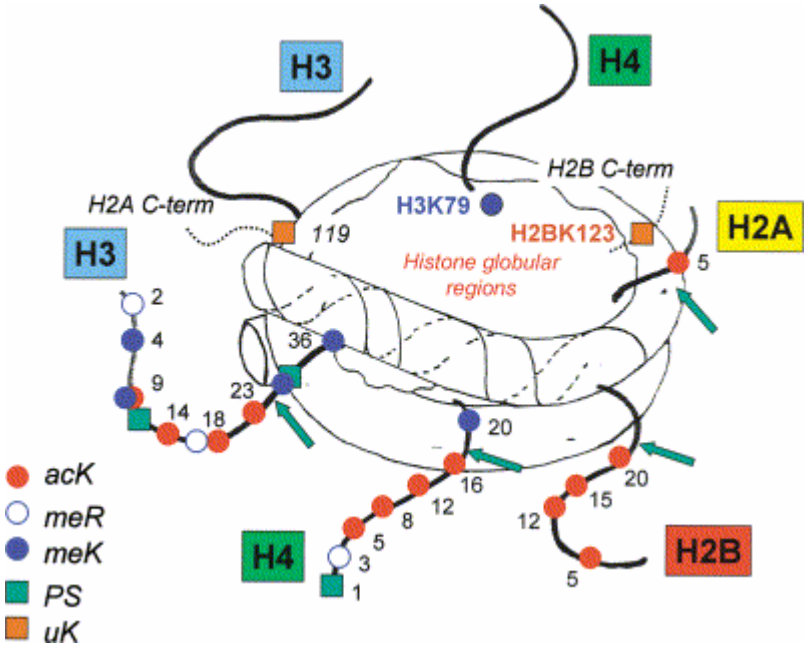

Figure 1.2: Histone modifications on the nucleosome core particle. Illustration of how the histone tails protrude out of the nucleosome core particle. Colored symbols indicate sites of post-translational modifications that are defined at the lower left; acK, acetyl lysine; meR, methyl arginine; meK, methyl lysine; PS, phosphoryl serine; uK, ubiquitinated lysine. Numbers indicate residue for each modification. Notice that H3-K9 can be both methylated and acetylated (Figure adapted from Turner (2002)

1.1.3 Histone modifications

One of the best understood modifications is acetylation, which in general is linked to transcriptional activation because histone acetylation decreases inter-nucleosome interaction and interaction of histone tails with linker DNA, thereby allowing greater accessibility (Khorasanizadeh 2004). The acetyl group can be removed by deacetylases which functions as repressors, thus, the combined actions of these enzymes serve as a rapid mean of switch between active and repressive states (Vermaak et al. 2003).

Although all histones are phosphorylated in vivo, the function of many of these modifications and the kinases that carry them out are not known (Isenberg 1979). Over the resent years some progress has been done in understanding what functions this mark can have on transcriptional regulation; phosphorylation of histone H2B at serine 14 (S14) in vertebrates is associated with apoptotic chromatin, whereas in all eukaryotes phosphorylation of histone H3-S10 is associated with transcriptionally

active and mitotic chromatin (Goll and Bestor 2002; Cheung et al. 2003; Fischle et al. 2003).

Histone methylation is catalyzed by histone methyl-transferases (HMTs), but it does not change the overall charge of the nucleosome. HMTs can be grouped into two families: protein arginine methyl-transferases (PRMTs) and histone lysine methyl-transferases (HKMTs). Two types of PRMTs transfer methyl groups to protein substrates; Type I PRMT enzymes form mono-methyl-arginine and asymmetric di-methyl-arginine products, while Type II PRMT enzymes catalyze the formation of mono-methyl-arginine and symmetric di-methyl-arginine (Gary and Clarke 1998; McBride and Silver 2001). Methylation of specific arginines on histones H3 and H4 correlate with the active state of transcription (Lee et al. 2005).

Histone lysine methylation has been considered to be a long-term epigenetic mark of maintaining chromatin states due to the fact that no demethylase had been identified (Jenuwein and Allis 2001). However, other mechanisms than enzymatic activity had been proposed for removal of1u histones methylated on lysines from the nucleosome (Jenuwein and Allis 2001; Ahmad and Henikoff 2002; Bannister et al. 2002). The resent identification of a histone lysine demethylase has therefore provided some surprise (Kubicek and Jenuwein 2004; Shi et al. 2004). This discovery, in addition to the putative revelation of other demethylases will not, however, necessarily change the notion that histone lysine methylation has a role in epigenetic regulation of genes (Kubicek and Jenuwein 2004).

1.1.4 SET-domain proteins

Genes encoding mammalian homologues of the *Drosophila* suppressor of PEV *Su(var)3-9* were the first to be shown to encode proteins with HKMT activity (Rea et al. 2000). A hallmark signature of this class of HKMTs is the presence of the 130-amino acid SET-domain, which is crucial for catalytic activity but also requires adjacent cysteine-rich domains (Rea et al. 2000). The evolutionary conserved SET-

domain found in numerous chromatin regulators, was named for its occurrence in genes encoding three such regulators in *Drosophila*, namely *Su(var)*, *E(z)* and *trithorax* (Jenuwein et al. 1998). Many SET-domain proteins have been shown to possess HKMT activity towards specific lysine residues on histone tails, leading to positive or negative regulation of gene expression (Kouzarides 2002; Lachner and Jenuwein 2002). Substrate specificity of SET-domain HKMTs is thought to be modulated through combination with the two flanking cystein-rich motifs called PRE-SET and POST-SET (Lachner and Jenuwein 2002). Methylated lysines are identified by proteins containing chromodomains, which were first identified in two regulators of chromatin structure in *Drosophila*: heterochromatin protein 1 (HP1) and Polycomb (Paro and Hogness 1991). So far, SET-containing HKMTs that methylate K4, K9, K27 and K36 of histone H3 and K20 of histone H4 have been identified (Cheng et al. 2005). Currently, all but one of the known histone residues that are methylated (H3-K79) are modified by SET-domain proteins (Feng et al. 2002; Min et al. 2003). The HKMTs have been reported to mono-, di- and trimethylate lysine residues, leading to a possible increased level of complexity in the interpretation of the histone code (Dutnall 2003).

73 entries have been found in the human genome which possess a SET-domain, 6 in *Saccharomyces cerevisiae*, 11 in *Schizosaccharomyces pombe*, 41 in *Drosophila* and 37 in *Caenorhabditis elegans* (Kouzarides 2002). By alignment of the SET-domains of these proteins, they have been classified into four subgroups; E(Z), TRX, ASH1 and SU(VAR)3-9 (Jenuwein et al. 1998).

The best characterized methylation sites are lysines 9 and 27 on histone H3; the first is associated with heterochromatin-mediated silencing (Jacobs et al. 2001) and the other with epigenetic repression (Cao et al. 2002). H4-K20 methylation has also been shown to be associated with silent chromatin and prevents acetylation on H4-K16, a mark for active chromatin in humans (Cao et al. 2002). Transcriptionally competent euchromatin, however, is in general methylated at three positions; H3-K4, H3-K36 and H3-K79 (Sims et al. 2003).

1.1.5 The SU(VAR)3-9 family

SUV39H1 and Suv39h1 was the first HKMT identified in human and mouse, respectively, showing specific activity towards H3-K9 (Rea et al. 2000). These proteins are enriched in heterochromatin (Aagaard et al. 1999) and accumulate transiently at centromeric positions during mitosis (Aagaard et al. 2000). The homologue SU(VAR)3-9 has been shown to be the main HKMT specific for H3-K9 in *Drosophila*. The methylation of H3-K9 is, however, influenced by pre-existing H3-S10 phosophorylation and H3-K9 acetylation, the first associated with chromosome condensation or segregation, the other with histone deposition (Strahl and Allis 2000). On the other hand, methylation of H3-K9 interferes with phosphorylation of H3-S10 that is dependent on the kinase Ipl1/aurora (Rea et al. 2000). An interaction between SU(VAR)3-9 and the histone deacteylase HDAC1 *in vivo* has also been reported in *Drosophila* (Czermin et al. 2001), possibly leading to the generation of a more permanently repressed chromatin structure within an otherwise more accessible, acetylated chromatin (Czermin et al. 2001).

Methylated H3-K9 creates a specific binding site for HP1 (Bannister et al. 2001; Jacobs et al. 2001; Lachner et al. 2001; Schotta et al. 2002). These observations have provided new insights into the mechanism of heterochromatin formation and maintenance. After primary association of SU(VAR)3-9 with heterochromatin, consecutive H3-K9 methylation will create binding sites for HP1 through its chromodomain (Bannister et al. 2001; Lachner et al. 2001). This will finally result in stable association of SU(VAR)3-9/HP1 complexes with heterochromatin and providing a feed-back loop for the methylation of adjacent nucleosomes (Jenuwein 2001). This has also been supported by observations in *Drosophila* (Schotta et al. 2002).

Another well characterized member of the SU(VAR)3-9 family is the mammalian G9a. This protein shows specific selectivity to H3-K9, but in addition it is capable of methylating K27 on the same histone (Tachibana et al. 2001). It is essential for

embryonic development in mouse, and it is thought to be involved in transcriptional silencing of developmentally regulated genes in euchromatic regions (Tachibana et al. 2002).

1.2 *Drosophila melanogaster* as a model organism

Drosophila melanogaster provides an excellent model system in which to study genetics and development, in addition to epigenetics. Thomas Hunt Morgan became interested in heredity in the beginning of the 20th century, after Gregor Mendel's work on genetic linkage was rediscovered, and together with his colleagues he published "*The Mechanism of Mendelian Heredity*" (Morgan et al. 1915). This was based on their work with *Drosophila melanogaster*. The reasons why he chose this species as an experimental organism was that it is small, has a short life cycle, produces a øarge amount of progeny and is easy to maintain in the laboratory. Additional advantages in working with this species have been discovered as well; there is no meiotic recombination in males, making it relatively easy to track chromosomes through generations. *Drosophila melanogaster* has only four chromosomes, which can be directly visualized in the larval salivary glands as the giant polytene chromsomes, and have provided a valuable tool in mapping genes. Furthermore, the *Drosophila* exoskeleton shows a wealth of external features such as bristles, compound eyes and body color, which can be affected by mutations and directly identified by investigation using a stereomicroscope. This has made it possible to link genotypic changes to phenotypic traits. After being studied in the laboratory for almost 100 years, a considerable number of techniques have been developed that have given *Drosophila* its role as one of the most important model organisms for genetic analysis (Rubin and Lewis 2000).

Model organisms are used for studying human diseases and improving our knowledge about life, without having to encounter many of the ethical issues following research on humans, or even mice. Even though flies and mammals

diverged around 700 million years ago, many of the relevant developmental processes are essentially conserved (Adams et al. 2000). *Drosophila melanogaster* has only around 15000 genes, and surprisingly many of these have clear homologues in humans (Friedman and Hughes 2001), including genes that underlie many genetic disorders, including cancer (Fortini et al. 2000). The possibility to perform large genetic screens in *Drosophila*, one of the most important hallmarks of the organism, provides an unbiased way to identify genes in a particular process and a great potential to dissect a specific gene function (St Johnston 2002).

1.2.1 The life cycle of Drosophila melanogaster

Drosophila belongs to the order of *Diptera*, two-winged insects, and the family *Drosophilidae*. It is a holometabolous insect, meaning that it has a pupal stage interposed between the larval and adult form. The fertilized egg is laid in nutritious food, and the embryonic development lasts for about one day, before it hatches as a larva. The larva passes through three stages, or *instars*, where it spends almost all its time eating and gaining size. After about 40 hours in the third instar stage it climbs to a dry and clean place where it molts into a pupae. Here it undergoes metamorphosis where a radical remodeling of the body takes place, developing into the adult fly, also called *imago*. The whole process from fertilization to the final hatching of the adult fly takes about 10 days at 25ºC (Fig. 1.3).

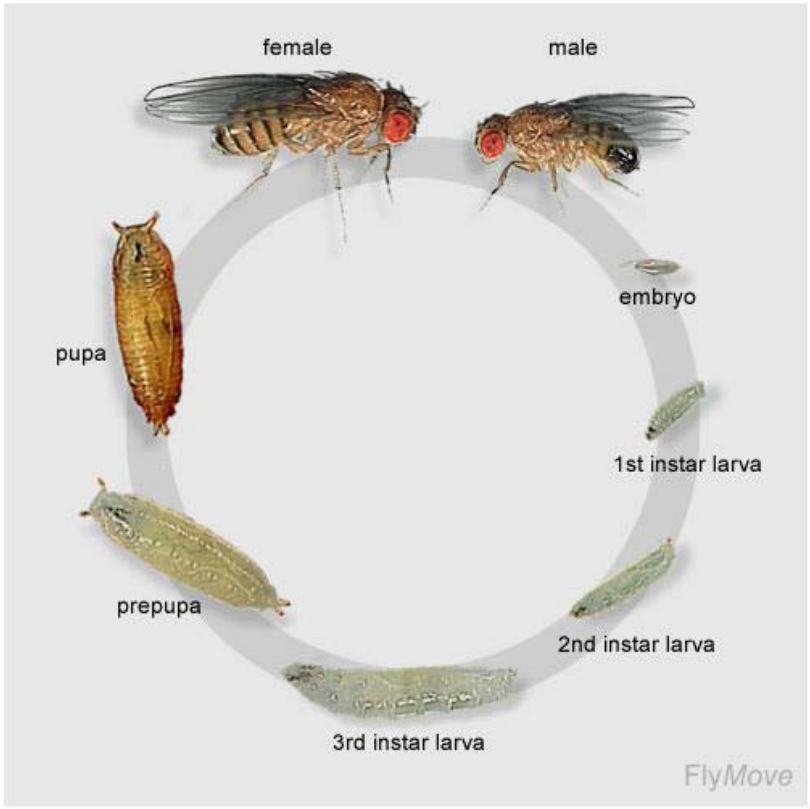

Figure 1.3: The life cycle of *Drosophila melanogaster*. The egg undergoes 1 day of embryogenesis, before it hatches into a larva. The first and second instars last for 1day each, before hatching into the last instar stage, which lasts for 2 days. The pupa goes through metamorphosis, and after 5 days the adult fly emerges. Image adapted from Fly*Move* (Weigmann et al. 2003).

1.2.2 Embryonic development in *Drosophila*

The embryonic development of *Drosophila melanogaster* lasts for one day, and has been subdivided into 17 stages (Campos-Ortega and Hartstein 1985). Embryogenesis starts with the cleavage during which the nucleus of the fertilized egg performs 13 rapid divisions; this takes place at stages 1-4. The divisions are not, however, followed by cytokinesis, resulting in a multinucleate cell called a syncytial blastoderm. At stage 5 the cellularization process starts, in which the resulting nuclei become arranged in a single layer beneath the egg surface of the embryo, and cell membranes are formed around the nuclei. The embryo is now called a cellular

blastoderm. At stage 6 the gastrulation process starts, leading to the formation of the three germ layers; ectoderm, mesoderm and endoderm. The subsequent development with various cell movements (e.g. germ band extension) results in an embryo with morphologically distinct segments along the anteroposterior axis.

1.2.3 The genetics of *Drosophila* embryonic development

Up to the cellular blastoderm stage development depends largely on maternal mRNAs and proteins that accumulate in the egg before fertilization, facilitating rapid development. These maternal-effect genes form the morphogenic gradients that are required to define the anteroposterior and dorsoventral axes of the embryo. They are transcription factors that regulate the expression of the zygotic gap genes, which roughly subdivide the embryo along the anteroposterior axis. Another level of complexity is added by the pair-rule genes; which are regulated by gap genes and divide the embryo into pairs of segments. Segment polarity genes are further controlled by pair-rule genes, and set the axis of each segment. All these classes of genes are together called segmentation genes, and they provide a hierarchy of signals that regulates segment patterning in *Drosophila* (Lawrence 1992*)*.

During embryonic development the segmentation genes are responsible for the activation of homeotic genes, which determines the identity of the different body segments along the anteroposterior axis (McGinnis and Krumlauf 1992). When the expression of the segmentation genes ceases shortly after gastrulation, the correct expression of the homeotic genes becomes dependent on proteins encoded by the *Polycomb*-group (*Pc*G) and the *trithorax*-group (*trx*G) genes. The PcG proteins keep homeotic genes repressed in cells where they need to be inactive, whereas the trxG proteins sustain expression of homeotic genes where their products are required (Orlando and Paro 1995; Pirrotta 1998). They regulate expression patterns at the chromatin level and provide a transcriptional memory mechanism (Simon and Tamkun 2002). The SET-domain proteins were initially identified as members of the

PcG and trxG families, implying that this group is involved in the regulation of early development of *Drosophila*.

1.2.4 Oogenesis

The early embryonic development is, as mentioned, supported by maternally deposited mRNAs and proteins into the unfertilized egg. Also, the main axes of the future fly body are defined before fertilization. These features are accomplished during oogenesis.

Drosophila ovaries are organized into approximately 15 ovarioles, each representing an independent egg assembly line. The ovarioles are tubular structures that contain progressively maturing egg chambers (Fig. 1.4), and are tipped with germ-line and somatic stem cells in which the development of the egg chambers is initiated. This region is called the germarium, and here each germ-line cell divides asymmetrically giving rice to one cytoblast and one stem cell. The cytoblast undergoes four rounds of cell division without completing cytokinesis, generating a cyst of 16 interconnected germ cells. One of these cells becomes the oocyte, while the remaining cells become nurse cells. Surrounding the 16 germ-line cells is a monolayer of somatic follicle cells. The subsequent development is divided into 14 morphological distinct stages (Spradling 1993). As oogenesis proceeds, the nurse cell nuclei become highly polyploid, whereas the oocyte remains arrested in meiotic prophase and is transcriptionally quiescent. Nurse cells actively transcribe and transport mRNA into the oocyte during stages 1 through 8 of oogenesis. At stages 9 through 10A a dramatic increase in nurse cell transcription is seen, and at stage 10B the nurse cells empty their remaining cytoplasm into the oocyte (dumping). This process is driven by a myosin-based contraction, and is followed by a modified form of cell death. The inducers of apoptosis in germ-line cells are not yet identified, although many of the molecules that execute the cell death program are likely to be the same as for somatic cells (Buszczak and Cooley 2000).

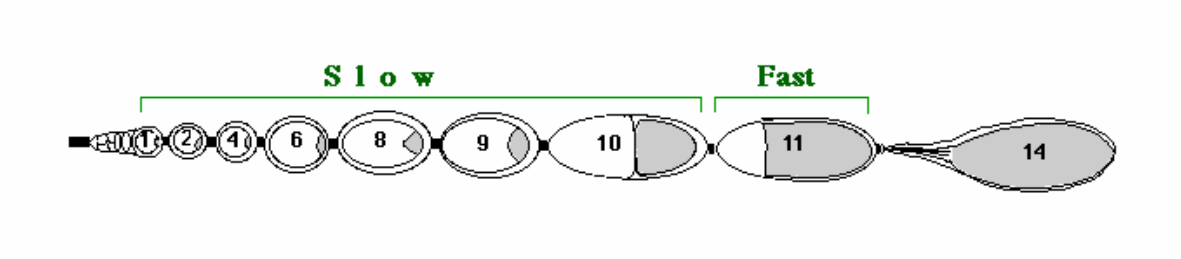

Figure 1.4: *Drosophila* egg chambers develop within ovarioles. Oogenesis proceeds in an assembly line manner through 14 morphological distinct stages. Figure adapted from Hudson and Cooley (2002).

The follicle cells surrounding the oocyte contribute with a complex set of signals that helps determine the future axis of the fertilized embryo, in addition to synthesizing the proteins of the tough eggshell (chorion) covering the egg chamber during the later stages of oogenesis (Margaritis et al. 1980).

1.2.5 Genetic tools for investigating gene function

In the beginning of the 20th century Thomas Hunt Morgan and his coworkers performed breeding experiments in their *fly room*. In 1910 their work was rewarded by the discovery of a mutant male fly with white eyes. Since spontaneous mutations occur very rarely, scientists have developed better ways to generate mutants, such as X-rays, gamma rays and chemical mutagens like ethyl methane sulphonate (EMS). In the beginning most mutagenesis studies were designed to generate point mutations in general. Christiane Nüsslein-Volhart and Eric Wieschaus were the first to perform a large-scale screen that set out to saturate the genome for mutations that affect a particular process, namely embryonic development (Nusslein-Volhard and Wieschaus 1980).

An additional way to generate mutations is by the use of transposable elements. The P-element was discovered in *Drosophila* as the causative agent of a genetic syndrome called hybrid dysgenesis (Engels 1983). The main advantage with P-elements is that they can be easily identified and mapped by sequencing (St Johnston 2002). However, the preferential insertion into 5´-noncoding regions makes it impossible to mutate every gene in the genome (Spradling et al. 1995). The Berkeley *Drosophila*

Genome Project supplied a large collection of P-element insertions to the Bloomington *Drosophila* stock center (Spradling et al. 1999), which are available for the public. P-elements also provide a popular tool for the creation of transgenic flies by P-element transformation, or for other purposes like enhancer-traps screens (Spradling et al. 1999).

1.2.6 The GAL4/UAS expression system

The GAL4/UAS system is one of the most elegant tools developed for targeted gene expression in *Drosophila*. It was developed by Brand and Perrimon (1993) and is based on the yeast transcriptional activator GAL4. This protein regulates the transcription of genes GAL10 and Gal1 in yeast by directly binding to four essential and related 17 bp sequences, called Upstream Activating Sequences (UAS) (Giniger et al. 1985). GAL4 expression is capable of stimulating transcription of any reporter gene under UAS control in *Drosophila* (Fischer et al. 1988), and the protein has no deleterious phenotypic effects, and does not activate native *Drosophila* genes. GAL4 transcription can be driven by various *Drosophila* promoters, resulting in expression in different tissues and at different times.

Several fly stocks have been developed, either with characterized *Drosophila* genomic enhancers or by enhancer detection techniques, where GAL4 is expressed in different patterns. Other stocks contain insertions with a polylinker located between four UAS sequences and a SV40 terminator. This polylinker makes it possible to ligate any gene or sequence you wish to express at a specific time or in a specific tissue (Fig. 1.5). The advantage of this system is that it is possible to generate fly strains with ectopic expression of the sequence at interest in different tissues or cell types. It also makes it possible to separate the target sequence from its transcriptional activator in two different strains where they have no effect, and only through a crossing between these two strains the target gene is turned on in the progeny, and it is possible to se a phenotypic effect (Brand and Perrimon 1993).

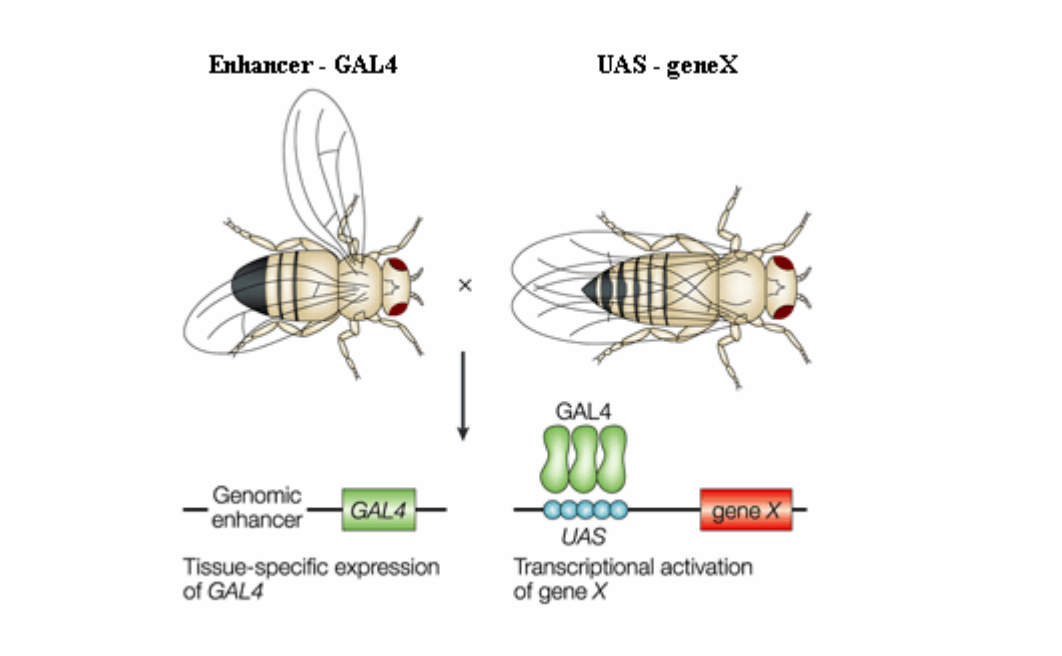

Figure 1.5: The GAL4-UAS expression system. The enhancer and target gene is brought together in the progeny by a single crossing. The expression of GAL4 is driven by a time and tissue specific enhancer, and regulates the expression of the gene of interest. Figure adapted from St Johnston (2002).

1.3 The Yeast Two-Hybrid system

The yeast two-hybrid (Y2H) system (Fig. 1.6) is a genetic method for identifying and analyzing protein-protein interactions. The principle of the assay was first published in 1989 (Fields and Song 1989), and has proven to be one of the most efficient techniques for finding new interactions (Fields and Song 1989; Frederickson 1998; Drees 1999). The method is based on the properties of the yeast GAL4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation respectively (Keegan et al. 1986). Plasmids encoding two-hybrid proteins are constructed and introduced into two different yeast strains of opposite mating type. One, pGBKT7 (BD Biosciences), consists of the GAL4 DNA binding (DNA-BD) domain fused to the protein of interest, the bait, whose interaction partners it is searched for. The other, pGADT7-Rec (BD Biosciences), consist of the GAL4 activation (DNA-AD) domain, fused to a cDNA library obtained form the species/tissue of interest. When the two yeast strains are mated, and a positive

interaction with the bait and a protein of the library occurs, this will lead to a transcriptional activation of four reporter genes; ADE2, HIS3, lacZ and MEL1. By streaking colonies on quadruple dropout medium (QDO) lacking these four reporter products, the appearance of false positives is reduced. This is termed a quadruple reporter system.

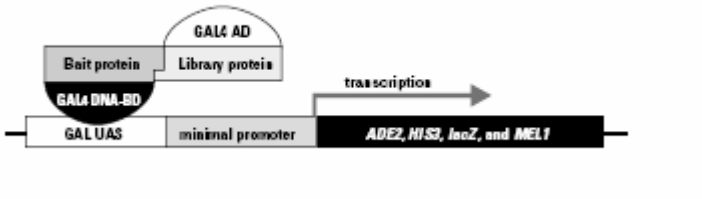

Figure 1.6: The yeast two-hybrid system principle. The bait is cloned into the DNA-BD vector where it is expressed as a fusion to the DNA binding domain of the yeast GAL4 protein. A second gene or cDNA library is cloned into the AD vector, where it is expressed as a fusion the activation domain of the GAL4. When the fusion proteins interact, the DNA-BD and AD domains are brought into close proximity and can activate transcription of reporter genes. (Figure adapted from the BD Biosciences Y2H user manual).

1.4 Aim of this project

This study is part of a larger project where the goal is to examine possible mechanisms of epigenetic regulation of transcription in *Drosophila melanogaster* and *Arabidopsis thaliana*. Although a considerably amount of work has been laid down in the investigation of the histone code hypothesis, it is a complex field of study, and much must still be accomplished. To get a better understanding of epigenetic regulation of transcription it is essential to characterize all genes that are involved in the process, in different organisms.

The aim of this thesis was to take advantage of genetic tools available in order to perform functional analysis of one putative SET-domain gene in *Drosophila*. The main objectives have been to:

- characterize the gene using bioinformatics tools
- investigate the *in vivo* localization and function
- identify putative interaction partners using the yeast two-hybrid system

2. Materials and methods

2.1 Fly stocks

2.1.1 Wild type

DmS, 93

2.1.2 Balancer stocks

yw; Sp/CyO; D/TM3,Sb

FM6,w

w^{1118}; If/Cyo

yw; D/TM3,Sb

yw; Sp/Cyo; Ki/TM3,Sb

2.1.3 Stocks used for the P-element excision

13414: y*[1] P{y[+mDint2] w[BR.E.BR]=SUPor-P}CG2995[KG01242*]

(http://rail.bio.indiana.edu/.bin/fbidq.html?FBst0013414&resultlist=fbstock19356.data)

Containing the element (*P{SUPor-P}*) inserted in the 5'-end of the gene *DmG9a*.

mus309^{D2}: *w^{1118}/yw; mus309^{D2}/TM3,Sb* and

Δ2-3, mus309^{D3}: *w^{1118}/yw; Sp/CyO; Δ2-3,mus309^{D2}/TM3,Sb*

mus309^{D2} and mus309^{D3} are two different alleles of mutations of the gene *IRBP* which is involved in double-strand DNA break repair and P-element transposition

Δ2-3 is the transposase catalyzing the reaction.

FM6,w

2.1.4 Stocks used for the over expression assay

#4414: y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]

(http://flybase.bio.indiana.edu/.bin/fbidq.html?FBst0004414&resultlist=fbstock27650.data)

GAL4 expression driven by the promoter of the ubiquitously expressed gene Act5c (http://flybase.bio.indiana.edu/.bin/fbidq.html?FBgn0000042)

#5460: w[*]; P{w[+mW.hs]=GAL4-da.G32}UH1

(http://flybase.bio.indiana.edu/.bin/fbidq.html?FBst0005460&resultlist=fbstock27960.data)

GAL4 expression driven by the promoter of the ubiquitously expressed gene *daughterless*. *(http://flybase.bio.indiana.edu/.bin/fbidq.html?FBgn0000413)*

2.2 Genetics

2.2.1 Over expression of the gene *DmG9a*

In order to investigate the function of the gene *DmG9a*, a construct was made for over expression of the gene. A PCR reaction was run using the primers 2995cDNAattB1 and cg2995Y2H4attB2, and pCRII-TOPO-*DmG9a* as template. The product of this reaction was Gateway cloned, via the vector pDONR/Zeo (Invitrogen Life Technologies), into the vector pPWF (http://www.ciwemb.edu/labs/murphy/Gateway%20vectors.html). This vector contains a strong UAS promoter, and a copy of the mini-white gene, in addition to a Gateway cassette. The vector was injected into w^{1118}

embryos (Ann Mari Voie) using P-element transformation. These transgenic flies were crossed to different balancer stocks in order to map what chromosome the insertion had happened. The resulting stocks were then crossed to #4414 or #5460 flies.

2.2.2 Promoter analysis

The *DmG9a* promoter was cloned into the pCaSpeR-AUG-ßgal (http://flybase.bio.indiana.edu/.bin/fbidq.html?FBmc0000224) vector, using standard cloning techniques. pCaSpeR-AUG-ßgal contains a *lacZ*-gene which the promoter was cloned in front of.

A PCR reaction was run on DNA isolated from wild type flies using the primers 2995prom/*Eco*RI and 2995prom/*Bam*HI, and DyNAzyme™ II DNA polymerase. This amplified the region between the ATG of *DmG9a*, and the ATG of the upstream gene CG3038. The product was isolated from agarose gel, and cloned into the pCR2.1-TOPO vector, using the TOPO cloning strategy. Both pCR2.1-TOPO-*DmG9a*prom and the pCaSpeR-AUG-ßgal vector were cut with the restriction enzymes *Bam*HI and *Eco*RI (Promega), and ligated to each other, using the T4 DNA Ligase (Promega), as described in the manufacturer's protocols. The vector was injected into w^{1118} embryos (Ann Mari Voie) using P-element transformation. These transgenic flies were crossed to different balancer stocks in order to map what chromosome the insertion had happened.

2.2.3 Preparation of DNA for injection

For injection of DNA into *Drosophila* embryos, 6 µg of cloned plasmid DNA and 2 µg of helper DNA (Δ2-3) were mixed together with 1/10 volume of 3M NaAc and 1.5 volumes of 96% ethanol. The DNA was centrifuged (2 min, 13000 rpm) and washed in 70% ethanol, before it was air-dried, and dissolved in 20 µl injection buffer (5 mM KCl, 0.1 mM phosphate buffer pH 6.8).

2.2.4 Generation of *DmG9a* mutants by P element excision

To learn about the *in vivo* function of the gene *DmG9a*, deletion mutants were tried created by imprecise excision of the P element (*P{SUPor-P}*) inserted in the 5'-end of the gene. The insertion itself has no effect on the transcription of the gene. In order to increase the yield of imprecise excision the *mus309* trans-heterozygous genotype were used. A *mus309* heteroallelic combination was used because of the linkage of each allele to a recessive lethal mutation(s).

Single, red eyed, non-TM3, Sb male flies red from the cross (*13414/w; mus309^{D2}/TM3, Sb X yw; Δ2-3, mus309^{D3}/TM3, Sb*) were crossed to FM6,w females. 190 crosses were set on 18^{o}C, and 70 on 25 ^{o}C. Female offspring with white eyes were crossed back with male FM6,w.

2.3 Drosophila methods

2.3.1 Dissection of ovaries

Ovaries from virgin, wild type flies were dissected in Ringer's solution (182 mM KCl, 46 mM NaCl, 3 mM $CaCl_2$, 10 mM Tris-HCl). For fixation the ovaries were agitated for 10 min in a mixture of 1 volume of fixation buffer (100 mM KH_2PO_4/K_2HPO_4 (~25:20) pH 6.8, 450 mM KCl, 150 mM NaCl and 20 mM $MgCl_2 \cdot 6H_2O$), 4 vol. of dH_2O, 1 vol. of 36% formaldehyde, saturated with heptane. The ovaries were devitellinized in methanol, and rinsed six times in ethanol, and stored at -20^{o}C.

2.3.2 Collection of embryos

Embryos from wild type flies were collected on apple juice plates (22.5 g agar, 750 ml H_2O, 25 g sucrose, 250 ml apple juice, 1.5 g nipagin dissolved in 3 ml ethanol) after 4 and 18 hours.

2.3.3 Fixation and devitellinization of embryos

Embryos used for immunostaining or whole mount *in situ* hybridization were dechorionated in 50% bleach for 2.5 min., and rinsed in dH_2O, alternating with brief washes of collection solution (0.04% Triton X-100, 0.7% NaCl). The embryos were transferred to a scintillation vial containing fixation buffer (1.3X PBS, 67 mM EGTA pH 8.0) together with formaldehyde (4%) and heptane (50%), and shook vigorously for 25 min. The lower aqueous phase was removed, and methanol added in order to remove the vitelline membrane. After 1 minute of shaking the devitellinized embryos would sink to the bottom in the lower methanol phase. These embryos were rinsed twice with methanol, and six times with ethanol, and stored at -20^{o}C.

2.3.4 *Whole mount in situ* hybridization

RNA whole mount *in situ* hybridization was performed essentially as described in Tautz and Pfeifle (1989); Jiang et al. (1991), the staining procedure was based on the DIG Nucleic Acid Detection Kit (Boehringer Mannheim GmbH).

Fixated and devitellinized embryos and ovaries were rinsed 4 times in ethanol, then rinsed with ethanol:xylene, with increasingly amounts of xylene, before they were rocked for 45 minutes in 1:1.25 ethanol:xylene. Ethanol was used for washing 5 times 3 minutes each, following 2 times 3 minutes rinses in methanol. The embryos and ovaries were then washed in PBT (1X PBS, 0.1% Tween 20), and incubated for 8 minutes in 4 µg Proteinase K in PBT. The digestion was stopped with 2 times rinsing followed by washing 4 times 2 minutes in PBT. The samples were refixed for 25 minutes in PBTF (5% formaldehyde in PBT). PBT was used for washing 5 times 2 minutes each, then together with hybridization solution (HS - 50% formamide, 5X SSC, 100µg/ml sonicated and boiled salmon sperm DNA, 50 µg/ml heparin, 0.1% Tween 20) in a 1:1 mixture for 10 minutes. The samples were washed for 2 minutes and prehybridized at 55^{o}C for ~1 hour in HS. 0.5 µl of the dig U antisense RNA probe was diluted in 50 µl HS, and denatured at 80^{o}C, before added to the ovaries and

embryos. Hybridization was performed for 18 hours at 55°C. A sense probe was used as negative control.

The samples were washed first for 1 hour, then 8 times for 15 minutes in HS. PBT and HS (1:1) were then used for washing 2 times 15 min, followed by 5 washes for 10 min in PBT. The anti-digoxygenin-AP antibody (Roche Diagnostics) was diluted 1:2000 in PBT, and incubated with the samples at 4°C over night (ON).

Embryos and ovaries were washed 4 times for 15 min in PBT, and then in staining buffer (SB -100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris-HCl pH 9.5, 0.1 % Tween 20) 2 times for 5 min. NBT/bCIP (Roche Diagnostics) were diluted 1:50 in SB, and added to the samples for staining for the amount of time needed to give appropriate amount of staining. PBT was used to stop the reaction by rinsing 3 times quickly. The samples were then rinsed in PBT with increasingly amounts of ethanol, the samples were inverted between each time, before they were rocked in 1:1.25 PBT:ethanol for 5 min. This was followed by 10 washes for 5-10 min each. Before mounting the samples were rinsed in quickly in xylene, before they were coverslipped in Permount (Fisher Scientific).

2.3.5 Probe for whole mount *in situ* hybridization

Probe for detection of DmG9a transcript was made by Marianne Stabell (unpublished results).

Probe for detection of the *tun* transcript was generated by running a PCR reaction with one of the positives from the Y2H screen, inserted into the library vector pGADT7-Rec as template. IS30084left and IS30084right was use for primers. The product was cloned into the vector pCR®II-TOPO by TOPO-cloning. 5 µg of the plasmid was cut with the restriction enzyme *Dde*I at 37°C for ~2 hours.

The DNA was extracted and diluted in DEPC-H_2O. 1 µl of linearized DNA were used as substrate for 15 U SP6 RNA polymerase (Promega), 10 µl of the reaction also

contained 1X transcription buffer (Promega), 1X DIG-labeled U NTP's (Roche Diagnostics GmbH), 5 mM of DTT (Promega), 48 U RNasin (Promega) and DEPC-H_2O. The reaction mix was incubated for 37°C for 2 hours. dH_2O was used to adjust the volume of the reaction to 25 µl, and 25 µl 2X carbonate buffer (120 mM Na_2CO_3, 80 mM $NaHCO_3$) was added, followed by incubation at 65°C for 10 min. 50 µl 0.2 M NaAc (pH 6.0), 10 µl 4 M LiCl, 1 µl tRNA (Sigma-Aldrich) and 300 µl ethanol was added, before freezing the reaction for 15 minutes at -20°C. The RNA was centrifuged (13000 rpm, 4°C) and washed in 70 % ethanol, before dissolved in 150 µl HS.

2.3.6 Immuno staining

Antibody staining was performed essentially as described by Robinson and Cooley (1997). Fixed and devitellinized embryos and ovaries were rinsed in 50:50 ethanol: PBS (130 mM NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO_4, pH 7.0), three times in PBS and washed three times for 10 minutes in PBS. The samples were then incubated for 10 minutes in PBT (1X PBS, 0.3 % Triton X-100, 0.5 % BSA). α-DmG9a (aa cc + 1623 -1637) was diluted 1:100 in PBT, and incubated ~18 hours at 4°C. PBT was used for rinsing 3 times, then washing 4 times for 15 minutes. Secondary antibody, donkey anti-goat conjugated with Cy3 (Jackson Laboratories), diluted 1:200 in PBT, and DAPI (1:100) was incubated with the samples for 2 hours at room temperature. PBT was used for rinsing and washing as before. Then PBS was used for rinsing 2 times. The samples was transferred to PBS:glycerol (1:1), and equilibrated for 20 minutes, before mounted on slide and cover slipped and concealed with nail polish.

2.3.7 Microscopy

All microscopy was carried out using a Zeiss Axioplan Imaging2 system equipped with Nomarski optics, epifluorescence attachment and cooled LSD imaging facilities.

2.4 Yeast two hybrid

The Yeast Two-Hybrid (Y2H) screen from BD Matchmaker™ Library Construction & Screening Kits (BD Biosciences) was used to identify proteins that DmG9a interacts with *in vivo*.

The baits were cloned using Gateway cloning into the modified vector pGBKT7-cassette B (Silja S. Amundsen, unpublished results), which was adapted to the Gateway® system. Two domains were created for screening; Domain1 and Domain2 They were created using primers cg2995Y2H3attB1/cg2995Y2H3attB2 and cg2995Y2H2attB1/cg2995Y2H3attB2 respectively.

The LiAc-method was used to the transformation of Y187 yeast cells with the pGBKT7-cassette B plasmid. Lithium acetate (LiAc) permeabilized the cell walls and DNA was added. Then the cells and DNA was co-precipitated with PEG and incubated in 30 min. at 30^{o}C. The cells were briefly heat-shocked, LiAc and PEG washed out by DMSO, and then plated on selective medium (SD/-Trp/X-α-Gal, SD/-His/-Trp/X-α-Gal and SD/-Ade/-Trp/X-α-Gal), to test the DNA-BD fusion for transcriptional activation. Inactive bait was indicated by white colonies on SD/-Trp/X-α-Gal, and by lack of growth on the other plates.

A colony containing inactive bait was tested for insert by running a PCR-reaction with gene specific primers. The same colony was used to make a bait-culture for testing the DNA-BD fusion for toxicity. The colony was inoculated in 100 ml SD/-Trp/Kan (20µl/ml) and grown ON at 30^{o}C with shaking to OD_{600} of 0.8. The cells were resuspended in 5 ml SD/-Trp, and then ready for mating.

The BD Matchmaker two-hybrid library was screened by yeast mating. The bait strain was mated with a previously made library host strain, in 47 ml 2X YPDA/Kan (50 µg/ml), by incubation at 30^{o}C ON with shaking at 40 rpm. Then the cells were resuspended to a total of 10 ml in 0.5 X YPDA/Kan (50 µg/ml), and plated on different dropout media. Mating efficiency was determined by plating on SD/-Leu,

SD/-Trp and SD/-Leu/-Trp in different dilutions. The rest was plated on TDO/X-α-Gal and QDO/X-α-Gal, to selection for yeast diploids expressing interacting proteins. TDO stands for triple dropout medium; SD/-His/-Leu/-Trp, and QDO stands for quadruple dropout medium; SD/-Ade/-His/-Leu/-Trp, and helps to prevent negative positives.

Positives on the QDO/X-α-Gal plates were re-streaked on SD/-Leu/-Trp/X-α-Gal 3 times to purify colonies containing only one plasmid. Then the positives were re-streaked on QDO/X-α-Gal to verify the phenotype.

After having verified the phenotypes yeast cultures of the positives were made, which was further prepped, and used for PCR reactions with the primers 5´AD LD and 3´AD LD Insert Screening Amplimers BD Biosciences. The PCR products were cut with the restriction enzyme *Alu*I to group the positives. Based on this grouping the samples were further sequenced to identify the genes (proteins) responsible for the interactions.

2.5 Protein methods

2.5.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

For detection of proteins, SDS-PAGE was used. The detergent sodium dodecyl sulfate (SDS) intercalates into hydrophobic parts of the protein and disrupts its compact folded structure. At the same time it coats the protein with a layer of negative charges which causes them to migrate towards the anode in an electrical field, regardless of the net intrinsic charge of the uncomplexed protein.

The SDS-PAGE gel consists of two layers. First a stacking gel (5 % acrylamide/Bis solution, 200 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.1 % APS, 0.1 % N, N, N', N'-Tetramethylethylenediamine) on top, and then a separating gel with various

concentrations of acrylamide (10/12 % acrylamide/Bis solution, 390 mM Tris-HCl pH 8.8, 0.1 % SDS, 0.1 % ammonium persulfate (APS), 0.04 % N, N, N', N'-Tetramethylethylenediamine). 5 µl sample buffer were added to 20 µl of the samples (60mM Tris-HCl (pH 6.8), 25 % glycerol, 2 % SDS, 14.4 mM 2-mercaptoethanol, 0.1 % bromphenol blue), and boiled at 95^{o} C for 5 min, followed by centrifugation (13.000) for 5 min. The gel was run in running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS) at 19-36 mA for 1 – 1.5 h in the electrophoresis apparatus. For size marker SeeBlue® Plus2 (Invitrogen Life Technologies) was used.

The gel was stained with Coomassie Brilliant Blue for 30 min, and either destained with boiling dH_2O for 20 and visualized, or destained (40 % methanol, 10 % acetic acid) for 1 – 1.5 h, rinsed twice (7 % methanol, 7 % acetic acid, 1 % glycerol and in dH_2O), dried and visualized. An untreated gel could also be used further in western blotting.

2.5.2 Recombinant protein expression in bacteria

In order to express protein in bacterial cells the vector pGEX-GAW-4E cassette B (Bitte Stenvik, unpublished results), containing a glutathione S-transferase (GST) tag as a fusion to the inserted sequence and a Gateway recombination cassette was used. The vector was electro transformed into *E. coli* BL.21 cells, which have high protein expression level.

To control the size of protein fragments used as bait in the Y2H screen, it was expressed and measured using a small scale protocol. Expression of protein was induced by adding 0.1 mM of the lactose analog isopropyl ß-D-thiogalactoside (IPTG) to cell culture (LB, 100 µg/ml ampicillin or 200 µg/ml carbenicillin), and incubating for 30 min at 28^{o} C. The size was measured by running both induced and uninduced samples on a standard 10 % SDS-PAGE gel.

2.5.3 Protein isolation

The GST fusion protein produced by the pGEX-GAW-4E cassette B vector can be extracted by using Glutathione Sepharose™ 4B (Amersham Biosciences) beads, which binds the GST-tag.

For protein isolation, larger amounts of cells was needed, and expression was induced in 250 ml cell culture (0.08 – 0.1 mM IPTG) for 2 h, and controlled by running a 10% SDS-PAGE gel comparing uninduced and induced samples. The cells were sedimented by centrifugation (4500 rpm) for 15 min at 4^{o} C, and resuspended in 20 ml cold Tris-HCl (50 mM). After a second sedimentation, the cells were resuspended in RIPA-buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM EDTA, 1% Nonidet® P 40, 0.5 % sodium deoxycholat, 1 tablet Complete protease inhibitor (Roche Diagnostics GmbH) per 50 ml buffer),with 0.025 % lysozyme, and incubated at room temperature for 15 min, before frozen down in liquid nitrogen. The cells were slowly thawed in running water, before they were disrupted by sonication (3 times 10 second at 50 % strength). The lysate was centrifuged (18.000 rpm) for 25 min at 4^{o} C. The supernatant was mixed together with 500 µl 50% solution of Glutathione Sepharose™ 4B beads (prepared according to the manufacturer's protocol), and put on a rotating wheel over night at 4^{o}C.

The beads were sedimented (3000 rpm for 5 min at 4^{o} C) and resuspended in 5 ml RIPA-buffer, and stored 10 min at 4^{o}C. This procedure was repeated twice, before finally resuspended in 250 µl RIPA-buffer. 50 µl of this mixture as boiled to release the protein, and run on a 10 % SDS-PAGE gel for control of the isolation.

2.5.4 *In vitro* translation

To translate the interaction partner found by the Y2H screen into protein, the TnT® Coupled Reticulocyte Lysate System (Promega) was used, incorporating ^{35}S- Met

(Montebello Diagnostics) into the protein. The reaction was set up as recommended by the manufacturer. The translation was controlled by running 1 µl on a 10% SDS-PAGE gel.

2.5.5 GST-pulldown assay

The GST-pulldown assay was used to verify interactions discovered by the Y2H-screen.

The *in vitro* translated ^{35}S-labeled protein was first precleared with Glutathione Sepharose™ 4B beads and NTN buffer (20 mMTris-HCl pH 7.5, 100 mM NaCl, 0.5 % Nonidet® P 40, 1 tablet Complete protease inhibitor per 50 ml buffer) for 45 min at 4^{o}C. The supernatant was either saved for use as input, or added to the isolated GST-fusion proteins, and incubated for 2 hours at 4^{o} C. After washing with NTN buffer, the protein was released from the beads by boiling, and run on a 12 % SDS-PAGE gel. After staining and destaining, the gel was dried, and signal detected using the Typhoon 9400 Variable Mode Imager (Amersham Biosciences).

2.6 Blotting methods

2.6.1 Southern blottting

Southern blotting was performed to detect changes in chromosomal DNA.

3 µg of genomic DNA from the excision flies were cut with the restriction enzyme *Cla*I at 37^{o}C over night, and ran on a 0.8 % (w/v) agarose gel (3 V/cm) (Sambrook and Russell 2001). Afterwards the gel was soaked in denaturing solution (0.5 M NaOH, 0.5 M NaCl) 2 times for 20 minutes, rinsed in dH_2O, and soaked in neutralization (0.5 M Tris-HCl pH 7.5, 1.0 M NaCl. Before blotting the gel was rinsed in 10X SSC (1.5 M NaCl, 150 mM Na-citrate).

Blotting

The DNA gel was blotted onto a Hybond-N^+ nylon membrane (Amersham Biosciences) over night, and baked at 65^{o}C for 15 minutes and UV-crosslinked as described by the manufacturer to bind the DNA covalently to the membrane (Hoefer® Scientific Instruments).

Probe labeling

Probes were made using the *redi*prime™ II kit (Amersham Biosciences) following the manufacturer's protocol. 25 ng of DNA were used for labeling with 1.5 µl of [^{32}P] dCTP (Montebello Diagnostics, 0.56 mBq). The incorporation reaction was incubated for 45 minutes at 37^o C.

The probe used for Southern blotting, was generated by using the primers 2995intron/biotin and 2995intron. DNA from wild type flies was used as template.

Hybridization

Hybridization with radioactive probes was essentially performed as described by Galau et al. (1986). All incubation steps, wetting (30 minutes), pre-hybridization (1.5 hours), hybridization (~18 hours) and washing (1 hour and 30 minutes) was performed at 68^{o} C, and all solutions contained 0.7X SSC (105 mM NaCl, 10.5 mM Na-citrate).

For detection, the membrane was wrapped in plastic and exposed for variable amounts of time in a cassette containing a Storage Phosphor Screen (Amersham Biosciences) and analyzed on a Typhoon 9400 Variable Mode Imager (Amersham Biosciences).

Stripping of membranes

In order to remove old probe, the membrane was incubated with 1 mM EDTA, pH 8.0 + 0.1 % SDS until the there was no radioactivity left.

2.6.2 Western blotting

Blotting

Western blotting was used to transfer proteins separated by SDS-PAGE to a solid support, to provide a starting point for immunodetection. The protein was transferred from the gel onto a PVDF-membrane (Immobilon™-P transfer membranes, Millipore), which in advance was given a hydrophilic treatment with methanol, then submerged in dH_2O before it was put in a blotting buffer. A sandwich complex consisting of supporting pads and filter paper (wetted in blotting buffer) was fitted into the gel holder and placed in cold blotting buffer (25 mM Tris, 192 mM glycine, 20 % methanol) and run at 100 V for 1 h, in the presence of a cooling element and a stirring magnet at 4^{o}C.

Immunodetection

The membrane produced by Western blotting was used for detection of specific proteins by antibodies. The blocking solution used consisted of PBS with 2.5% Skim-milk and 0.1% Tween 20. The membrane was placed in a suitable dish and incubated on a shaker in four steps. In the first step the membrane was incubated with 15 ml blocking buffer and the primary antibody (diluted 100x) for 90 min. This step was followed by washing; 5 x 5 min in 15 ml blocking buffer. In the third step 15 ml blocking buffer and 1.5 µl anti-rabbit horseradish peroxidase was incubated for 60 min. This was followed by a new series of washing. For detection, a chemiluminescent kit (SuperSignal Chemiluminescent Substrate; Pierce) was used, and the work performed in a dark room. An equal volume of the two substrates in the kit was mixed as described by the manufacturer. The membrane was incubated in this solution for 2 min before it was removed and the membrane was wrapped in plastic foil. Then an X-ray film (Kodak) was placed against the blot and exposed for various time depending on the amount of signal expected. The film was then developed in an Optimax Film Processor (Protec).

2.7 RNA/DNA methods

2.7.1 RNA isolation

Total RNA was isolated from 10-50 mg wild type adult flies, embryos and ovaries, using the Trizol® reagent (Invitrogen Life Technologies) according to the manufacturer's instructions, and quantified on a GeneQuant spectrophotometer (Pharmacia).

2.7.2 RT-PCR

cDNA was generated using SuperScript™III First-strand Synthesis System for RT-PCR (Invitrogen Life Technologies). The reactions were in general set up as described by the manufacturer. 5 µg of isolated RNA was incubated at 65° C for 5 min together with 0.78 mM dNTP's and 385 ng/µl pd$(N)_6$ (Amersham Biosciences) in a total reaction volume of 13 µl. After a brief centrifugation the following reagents was added; 4 µl 5X first-strand buffer, 1 µl 0.1 M DTT and 1 µl SuperScript™III (50 units). The samples were then incubated at 50°C for 60 min before the enzyme was inactivated at 70°C for 15 min. The cDNA could now be used as template for amplification by PCR; 1 or 2 µl of the first-strand reactions were used. The PCR reactions were set up with the DyNAzyme™ II DNA Polymerase (Finnzymes) using standard conditions as described in chapter 2.8.1.

2.7.3 DNA from single flies

A single fly was homogenized with a pipette tip containing 50 µl Squishing buffer (10 Mm Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K) for 10-20 seconds. The buffer was then emptied in the tube, which was incubated at room temperature for 30 minutes. The Proteinase K was inactivated by incubation for 1-2 minutes at 95°C. The DNA was stored at -20°C.

2.7.4 CTAB-protocol

0.1-0.5 g flies were incubated with 2X CTAB (2% CTAB, 1 M Tris-HCl pH8, 0.5 M EDTA, 5M M NaCl diluted in 2-mercaptoethanol) and 100 µg Proteinase K, at 60°C for 1 hour. The DNA was extracted in 1X phenol:chloroform (1:1) and 1X chloroform, the samples were inverted on wheel for 15 min and centrifuged (13.000 rpm) for 20 min each time. The top phase was transferred to a new tube, and 96% ethanol was added, the sample was then incubated at RT for 10 min, before

centrifugation (13.000 rpm) for 5 min. 70% ethanol was used for washing the pellet, which was dried and dissolved in 50 µl dH_2O.

2.8 Cloning

2.8.1 PCR

Standard PCR reactions were set up with DyNAzyme™ II DNA Polymerase (Finnzymes) or PuRe Taq Ready-To-Go™ PCR beads (Amersham Biosciences). For longer amplifications BD Advantage™ 2 Polymerase (BD Biosciences Clontech) or Platinum® Taq High Fidelity (Invitrogen Life Technologies) was used. In the reactions where high-fidelity amplification was required, *Pfu* DNA Polymerase (Fermentas Life Sciences) was preferred, because of its 3'=>5' exonuclease (proofreading) activity.

200 µM dNTP's was already included in PuRe Taq Ready-To-Go™ PCR beads; in all other reactions 160 µM was used. Primers (Invitrogen Life Technologies) were always diluted to a final concentration of 160 nM each. Annealing temperature varied according to the melting temperature (T_m) of the primers in the reaction, usually between 50-60^o C. For reactions with BD Advantage™ 2 Polymerase, both annealing and extension was performed at 68^o C. Extension time varied with the length of the expected PCR product, and number of cycles was normally 25, but was sometimes increased to 30 or 35, to obtain more product. All other parameters followed the manufacturers' recommendations. The specific primers used can be found in Appendix 1.

2.8.2 Agarose gel electrophoresis

DNA were size fractioned on 1.0-1.2 % agarose gel's (SeaKem® LE agarose, Cambrex Biosciences, 1 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA), 0.6 μg/ml ethidium bromide). The markers GeneRuler™ 1kb DNA ladder (Fermentas Life Sciences) and φχ174 Hae III (Promega) were used for comparison. The gel was run at 60-100 V in 1 x TAE buffer for 30 to 60 min.

For DNA purification from gel the Wizard® *Plus* Minipreps DNA purification system (Promega) was used, following the supplier's protocol.

2.8.3 Ligations, restriction digests and transformations

Ligations and digests were set up as specified by the manufacturers, using T4-ligase from Promega and restriction enzymes from Fermentas, New England Biolabs or Promega.

Plasmids were transformed into either chemically Library Efficiency® DH5α Competent Cells (Invitrogen Life Technologies), or electrically competent BL21 *E. coli* cells, using standard techniques (Sambrook and Russell 2001). Bacteria was grown on standard LA-agar plates or in LA medium (Sambrook and Russell 2001), containing different antibiotics depending on what resistance gene the vector contained.

2.8.4 DNA purification

Plasmid DNA was purified from bacteria using either Wizard® *Plus* SV Minipreps (Promega), Quantum Prep® Plasmid Miniprep Kit (Bio-Rad Laboratories) or Qiagen® Plasmid Maxi Kit (Qiagen), following protocol from supplier.

2.8.5 DNA quantification

DNA was quantified using the Hoefer DyNAQuant 200 fluorometer (Hoefer® Scientific Instruments) using the fluorescent dye Hoechst 33258, according to manufacturer instructions.

2.8.6 Sequencing

Sequencing were performed at the Department of Molecular Biosciences (IMBV) sequencing facility, which uses MegaBACE™ 1000 and DYEnamic ET Dye Terminator Cycle Sequencing Kit (Amersham Biosciences)

2.8.7 Gateway cloning

The Gateway® cloning system (Invitrogen Life Technologies) was used to generate vectors used in various assays. This system is based on the site-specific recombination properties of the bacteriophage λ, which are conservative and highly specific, when it is integrated into the bacterial genome in the switch from lysogenic to lytic life cycle. The recombination involves two major components; the DNA recombination sequences (*att*-sites), and the proteins that mediate the recombination reaction. The *att*-sites serves as binding sites for recombination proteins (Weisberg et al. 1983).

The fragment to be cloned was amplified by PCR with primers containing two different *att* sequences, *att*B1 and *att*B2, to maintain orientation of the fragment. In the BP reaction the PCR-product was recombined into entry vector by, which further was cloned into a destination vector by the LR reaction. All reactions were set up as described in the manufacturer's protocol.

2.8.8 TOPO cloning

Some DNA polymerases add a deoxyadenosine (A) to the 3' end of the PCR products. These ends align with the overhanging 3'-thymidine (T) on the TOPO-vectors. A Topoisomerase I, which is normally found in the *Vaccinia* virus, is covalently bound to this vector, and uses these overhanging ends to ligate the PCR product into the vector. The reactions were set up as described by the manufacturer (Invitrogen Life Technologies).

2.9 Bioinformatics

Database searches for protein sequences were performed using BLASTP whereas for nucleotides the BLASTN database was used (http://www.ncbi.nlm.nih.gov/BLAST/). For identification of protein domains the SMART database was used (http://smart.embl-heidelberg.de/, (Schultz et al. 1998)). Nuclear localization signals were identified by running a search in the PredictNLS online database (http://cubic.bioc.columbia.edu/cgi/var/nair/resonline.pl). Multiple alignments of protein sequences were done with the ClustalX program (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/(Thompson et al. 1997)) and manually adjusted with the GeneDoc program (http://www.psc.edu/biomed/genedoc/). For DNA sequence analysis and primer analysis the Vector NTI v 9.0.0 (Informax).

3. Results

3.1 DmG9a is a putative histone methyl-transferase

In order to identify genes involved in epigenetic regulation of transcription, the "signature motif" of lysine-specific HKMTs, the SET domain, was used as bait in a BLASTP search. The search was performed using the SET domain of the well characterized *Drosophila Su(var)3-9* gene, and identified the annotated CG2995 as a putative HKMT gene in *Drosophila melanogaster* (Marianne Stabell, unpublished results). Based on the homology of the SET domain, the proteins in this family is divided into four families, SU(VAR)3-9, E(Z), ASH1, and TRITHORAX (Jenuwein et al. 1998), and CG2995 belongs to the first family. By using the SMART database (Schultz et al. 1998), one cysteine rich preSET domain, 6 ankyrin repeats and a coiled coil domain were identified, in addition to the conserved SET domain (Fig. 3.1A). A BLASTP search using the whole sequence of CG2995 protein showed that it is homologous to mammalian G9a with 30% amino acid identity and 47% amino acid similarity. G9a is a characterized HKMT , and shares the same modular design as CG2995, with a preSET domain, 6 ankyrin repeats, a coiled coil domain and, in addition, a cysteine rich postSET domain (Fig.3.1B; (Milner and Campbell 1993). Therefore, CG2995 will go under the name *Drosophila melanogaster G9a* (*DmG9a*). Alignment of DmG9a, the *Drosophila* SU(VAR)3-9 and *Mus musculus* G9a showed that DmG9a has the four sequence motifs thought to involved in the methylation process (Cheng et al. 2005) in addition to the conserved postSET domain. (Fig. 3.1C). *DmG9a* located on the X-chromosome and has been mapped to cytological band 1A1 (Fig. 3.1D) (The Fly*Base* Consortium, 2003). Based on the fact that DmG9a shows homology to the mammalian SET domain encoding gene G9a, which has been shown to have methyl-transferase activity, it was chosen for further studies in *Drosophila melanogaster*.

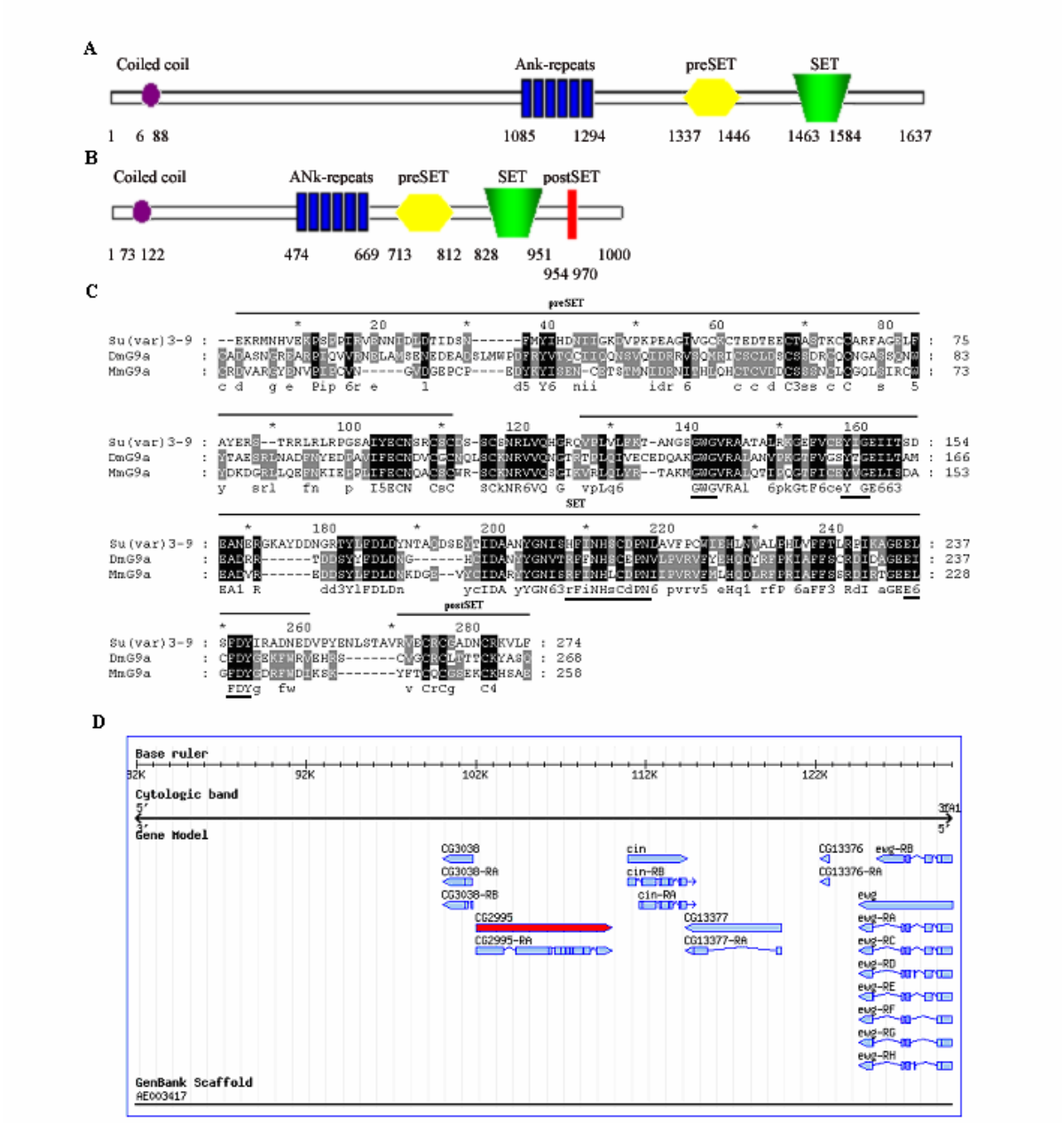

Figure 3.1: *Drosophila DmG9a*: Identified domains, alignment with homologues and genomic location. (A) Modular design of DmG9a based on domains identified by the SMART database. Numbers below indicate amino acids. (B) Modular design of MmG9a. (C) Alignment of DmG9a (CG2995), *Drosophila* SU(VAR)3-9 (AE003708) and MmG9a (AAC84164). The degree of conservation is distinguished at four levels (100, 80 and 60%, and not conserved), where 100% has the darkest shade of gray. The upper and lower case letters in the consensus line indicate 100 and 80% conservation within all groups, respectively. Only 100%, 80% and not conserved was identified. Numbers in the consensus line represent conserved similarity groups as defined by the Blosum 62 scoring table. Lines directly below the consensus sequence indicate the four conserved motifs revealed by (Cheng et al. 2005). (D) *DmG9a* (CG2995) highlighted in red, is located on the X-chromosome, cytological band 1A1. The figure shows other genes located close to Dm*G9a*, but there are no other genes predicted to overlap with it (adapted from Fly*Base*; The Fly*Base* Consortium, 2003).

3.2 Investigating the *in vivo* function of *DmG9a*

An effective way of dissecting the function of a gene is to create a knock-out mutant which may indicate which processes it is involved in, and whether it is essential or redundant. In *Drosophila*, one popular strategy is to use P-elements inserted in the coding or regulatory region of the gene in question. For *DmG9a*, stock #13414 from the Berkeley *Drosophila* Genome Project was available at the Bloomington Drosophila stock center (Spradling et al. 1999). This stock has the element *P{SUPor-P}* inserted into the 5'-UTR. This insertion, however, do not cause any phenotypic effects, and the flies have normal expression of *DmG9a*. This was shown by Southern hybridization and RT-PCR (Marianne Stabell, unpublished results). By mobilizing the P-element to transpose imprecisely, we hoped to create deficiencies in the *DmG9a* gene and thus a reduction of *DmG9a* transcript. To mobilize the *P{SUPor-P}*element the transposase Δ2-3 was crossed into the #13414 line, together with the two *mus309* alleles in trans, which increases the rate of imprecise excision. 260 crosses were set up and the progeny was screened for white eyes. 19 lines with possible excisions were found. These were further investigated by Southern and PCR to map the excision, but did not give any usable results.

The assay was based on the belief that the P-element had the eye-color marker w^+ (mini-white), which gives red eyes, and identifies an insertion against a background of non-transformed flies with white eyes. It was later found that it was instead the body-color marker y^1, which was used on this specific P-element. This would explain the difficulties of mapping the excision direction and size. Since the assay was performed on the wrong basis, it would be necessary to perform new crosses in order to obtain new stocks that could be used for screening of mutants. Due to lack of time this was not possible.

Another approach to investigate *in vivo* function is to express the gene of interest ectopically. There are a large number of P-element vectors for *Drosophila* based on the GAL4/UAS system where one could clone the gene of interest behind a UAS

promoter, which can be activated by Gal4. One of these vectors is pPWF, which is also adapted to the Gateway-system. *DmG9a* was successfully cloned into pPWF and injected into *Drosophila* embryos. Three lines of transgenic flies were obtained and crossed with flies which had Gal4 ectopically expressed. The progeny did not show any phenotypic changes compared to wild type flies. Virgin female flies from the F_1 progeny were crossed to wild type males in order to investigate if ectopic expression of *DmG9a* affected oogenesis. However, this cross did not give any phenotypic abnormalities.

Admittedly, over-expression of DmG9a protein was not confirmed, so it is unknown whether the lack of phenotype was due to redundancy of the DmG9a protein itself, or that the over-expression vector failed to be expressed at all.

3.3 *DmG9a* mRNA is present in embryos and ovaries of *Drosophila*

Whole mount *in situ* hybridization was performed on *Drosophila* embryos and ovaries in order to detect the transcript of *DmG9a.*

At stage 1-4 of embryo development a series of 13 mitotic nuclear divisions, unaccompanied by cytokinesis, takes place to produce a multinucleated single cell embryo, the syncytial blastoderm. Moderate levels of *DmG9a* mRNA were uniformly spread along embryos at this stage (Fig. 3.2A). This was also seen in the blastoderm of stage 5 embryos (Fig. 3.2B), which is when cellularization starts and a multicellular embryo is formed. These results indicate that the gene is maternally supplied, because there is little or no transcription before stage 5 in *Drosophila* embryos. During gastrulation *DmG9a* transcript was seen in the area where ectoderm and mesoderm develops (Figs. 3.2C and D). Mesoderm is the precursor cells of the fat body, muscles, heart and the somatic portion of the gonads, while ectoderm gives rise to the epidermis, tracheal tree, nervous system and the fore- and hindgut and their

annexes. The negative control with *DmG9a* sense probe did not show any staining that would interfere with the interpretation of the results (Fig. 3.2E).

Whole mount *in situ* hybridization using an antisense probe specific for *DmG9a* indicated that the gene is maternally deposited in the oocyte, and is expressed in the progenitor cells of various tissues in the developing embryo.

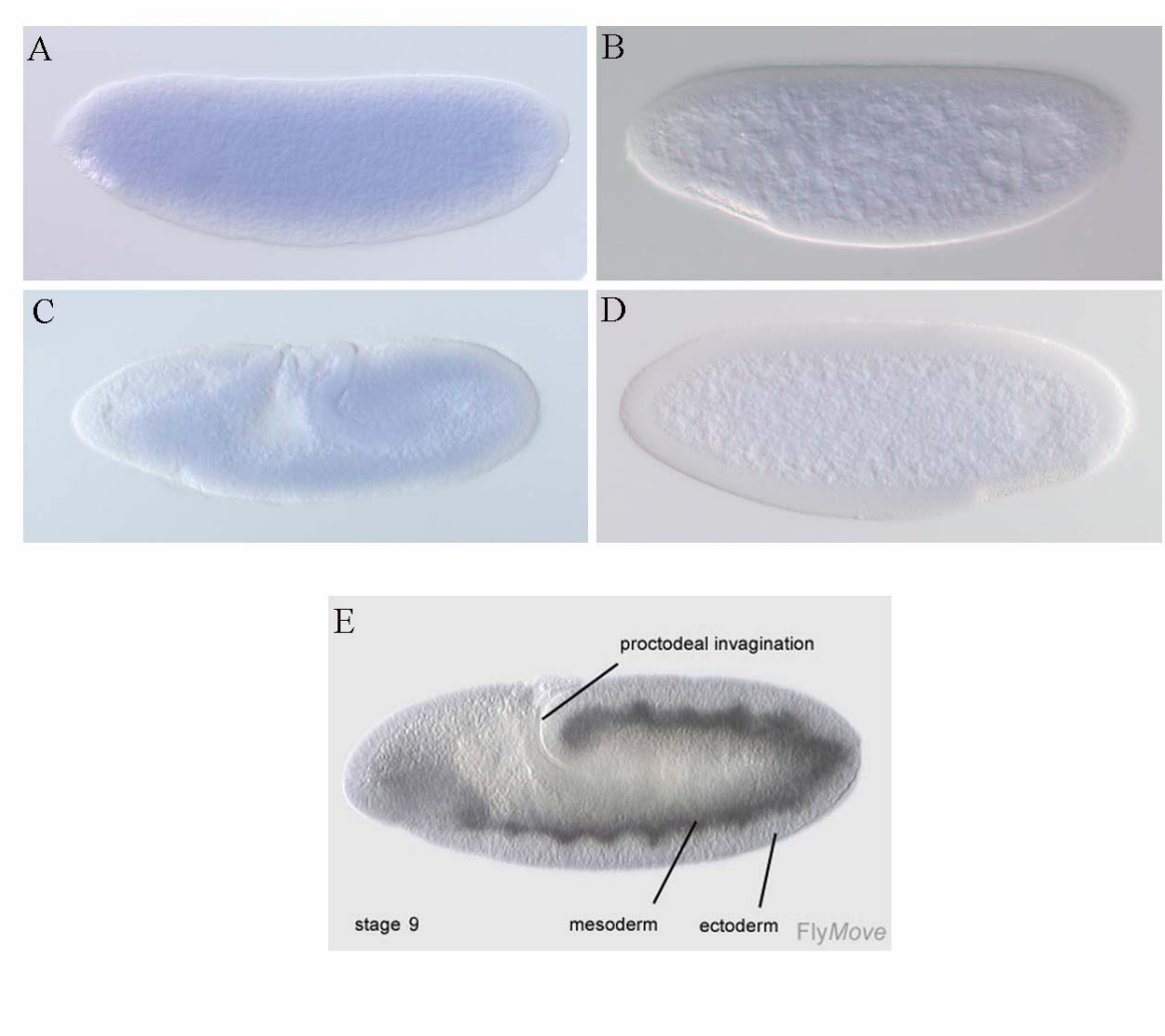

Figure 3.2: Whole mount *in situ* hybridization of *Drosophila* embryos against *DmG9a*. (A) Stage 4 embryo, a syncytial blastoderm. (B) Stage 5 embryo; cellularization starts. (C) Stage 8 embryo during gastrulation. (D) A stage 5 embryo hybridized with a *DmG9a* sense probe for negative control. (E) Showing a stage 9 embryo stained with antibody against the gene *twist*, which is expressed within the entire mesoderm Picture adapted from Fly*Move (Weigmann et al. 2003)*

Nurse cells synthesize and transport the cytoplasmic components that support the development of the oocyte and, later, the embryo. During the stages 1-9 of oogenesis,

the nurse cells produce moderate amounts of transcripts, which are transported to the oocyte. At stage 9 through 10A a dramatic increase in transcription starts, which is seen by the increase in oocyte size. *DmG9a* transcript was clearly visible in the oocytes from stage 9 (Fig. 3.3A), and weakly in nurse cells. At stage 10 the transcript was strongly expressed both the cytoplasm of nurse cells and in the oocyte (Figs. 3.3B and C). Towards the final steps of oogenesis there is no or very little *DmG9a* transcript in the oocyte, whereas it still remains in the nurse cells (Fig. 3.3D). In the mature oocyte there is no detectable *DmG9a* transcript (Fig. 3.3E).

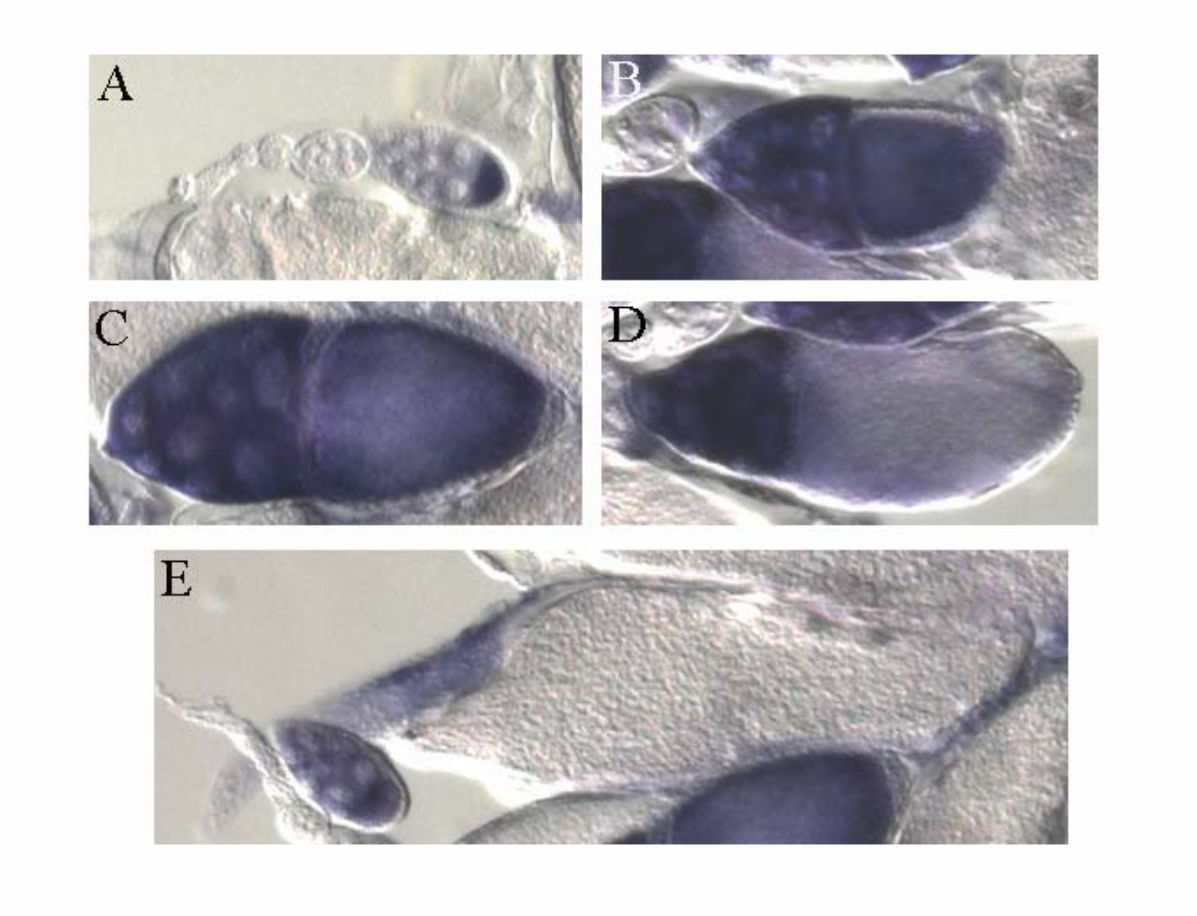

Figure 3.3: Whole mount i*n situ* hybridization of *DmG9a* in ovaries. (A) Stages 1-9. **(B)** Stage 10A. **(C)** Stage 10 B. A microfilament-dependent contraction of the nurse cells begins now. **(D)** Stage 11. Nurse cells deposit their cytoplasmic constituents into the oocyte before the nurse cells degenerate during stage 12. **(E)** Stage 14. Egg ready for fertilization.

3.4 DmG9a protein is present in nuclei of embryos, and both nuclei and cytoplasm of ovaries

DmG9a encodes a putative histone methyl-transferase, and the protein was therefore expected to be located in the nucleus. In line with this notion a putative nuclear localization signal, *PKRSKRR,* was identified when the DmG9a protein sequence was submitted into the PredictNLS online database (http://cubic.bioc.columbia.edu/cgi/var/nair/resonline.pl).

As Figure 3.4 shows the nuclear localization was also confirmed by antibody staining against DmG9a in *Drosophila* embryos. DmG9a is associated with nuclei in the syncytial blastoderm at stage 4 (Fig. 3.4A). At stage 5 DmG9a is also detected in the uncellularized nuclei in the center of the embryo, in addition to the blastoderm cells around the perimeter of the entire egg (Fig. 3.4C). After gastrulation (stage 9) DmG9a protein could be seen in the ectoderm (Fig. 3.4E), which gives rise to epidermis and nervous system, etc.

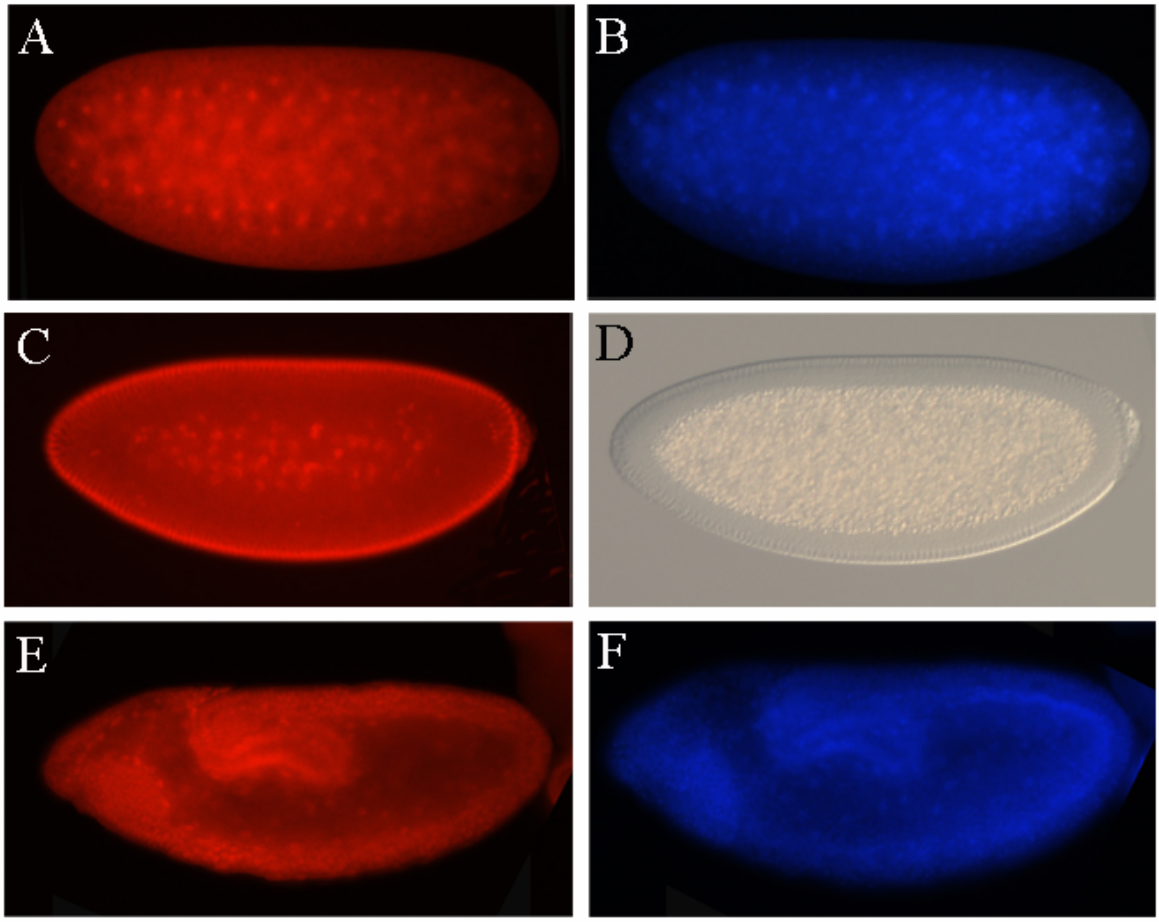

Figure 3.4: DmG9a expression during *Drosophila* embryo development. (A and B) Stage 4 embryo (A) antibody against DmG9a and (B) DAPI staining. (C and D) Stage 5 embryo, (C) antibody staining, (D) Nomarski (E and F) Stage 9 embryo, (E) antibody staining, (F) DAPI staining.

As Figure 3.5 shows, DmG9a protein also is present in ovaries of *Drosophila melanogaster*. It is detected in follicle cells from the early stages of oogenesis, but towards the later stages the protein level seems to be reduced (Figs. 3.5A-D). The somatic follicle cells participate in signaling which determines the oocyte, and later the polarity of the embryo. In addition, polarized follicle cells secrete the proteins making up the oocyte vitelline membrane and eggshell late in oogenesis.

DmG9a seems also to be present in nurse cell nuclei in the early stages of oogenesis, with the strongest signal around stage 10A (Fig. 3.5E). At the beginning of stage 11 the protein is shifted from the nuclei to the cytoplasm (Fig. 3.5G). This happens after a dramatic rearrangement of filamentous actin to form actin-bundles, approximately at the same time as the cell nuclear envelope becomes permeabilized. This means that

DmG9a is transported into the cytoplasm during initiation of apoptosis. DmG9a levels are reduced towards stage 14, along with the degeneration of nurse cells (Fig. 3.5I).

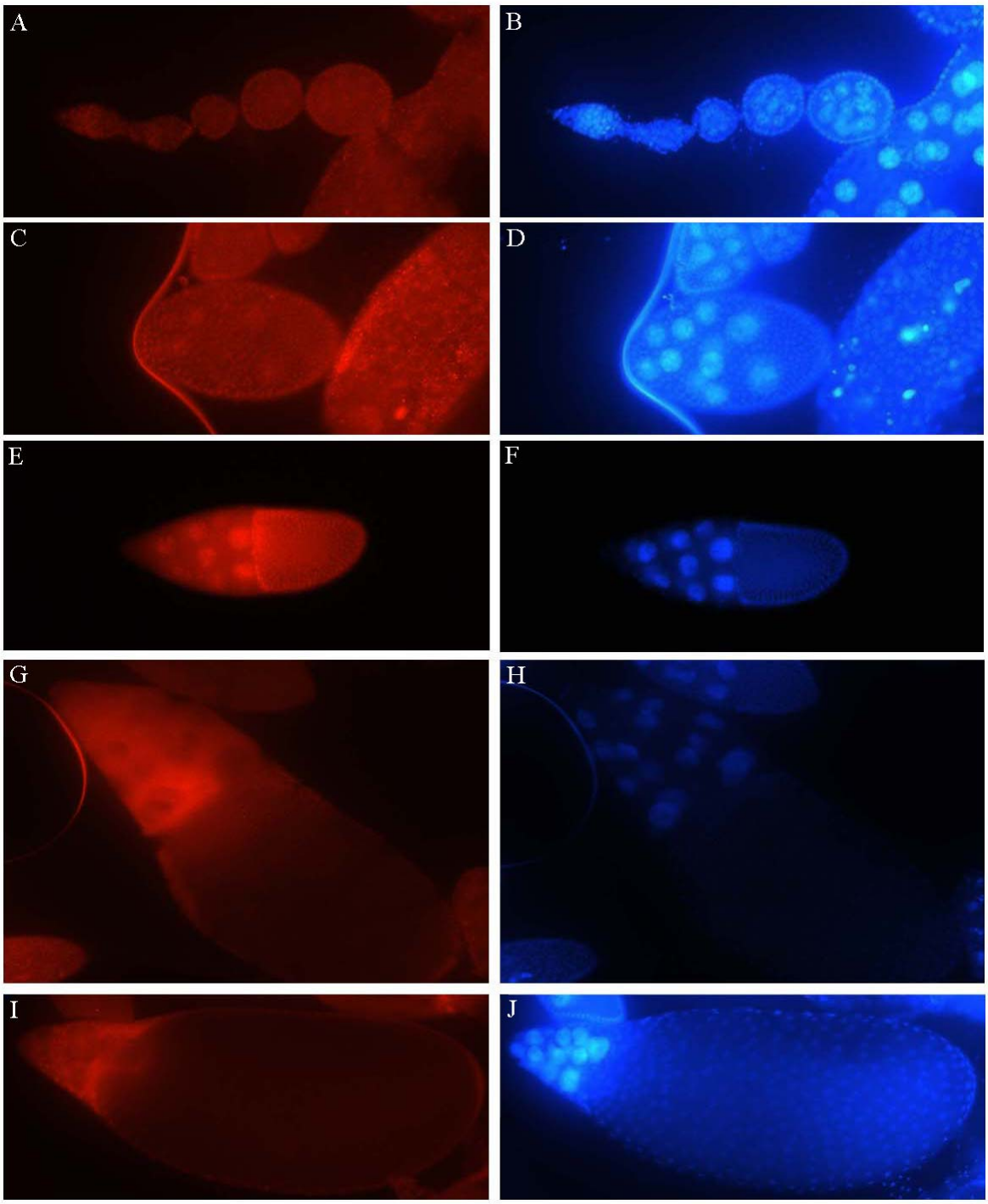

Figure 3.5: Immunostaining against DmG9a in ovaries. (A-B) Stage 1-8 egg-chambers. (C-D) Stage 9 egg-chamber. (E-F) Stage 10A egg-chamber. (G-H) Stage 12 egg-chamber. (I-J) Stage 11 egg-chamber. (A-I) α-DmG9a. (B-J) DAPI.

3.5 DmG9a interacts with the LIM-domains of Tungus

In this study the Y2H assay was used to analyze the protein-protein interaction profile of the DmG9a protein. Because of its relatively long ORF, which could be difficult to express in yeast, it was divided into smaller fragments, each of which was screened separately. To reduce the risk for the domains to not fold properly, whole modules was used for bait.

First, Domain 1 (Fig. 3.6 A), was used as bait. This fragment, from amino acids 1046 to 1335 contained the six ankyrin repeats (ANK), a ubiquitous motif which has been showed to function in protein-protein interactions (Bork 1993). The sequence was successfully cloned into the vector pGBKT7-cassette B and transformed into Y187 (MATα) yeast cells. This bait was found to autoactivate in these cells, which was seen by blue colonies appearing on SD/-His/-Trp and SD/-Ade/-Trp with x-α-galactocidase, before mating. The same was reported to happen with the DmG9a homologue HsG9a in a high-throughput yeast two hybrid screen performed by another group (Lehner et al. 2004). This is an indication of the domains' activity as transcriptional activator. The baits were expressed as fusions to the GAL4 DNA binding domain, and activated the transcription of the lacZ-gene, without help of the GAL4 activation domain. Acidic amphipathic domains are often responsible for unwanted transcriptional activation, and the pI of Domain 1 of DmG9a was calculated to 4.58. This bait, with its low pI was not suited for further research by the Y2H-system.

To circumvent the problem above, the bait was expanded and the more upstream region of the gene was included; resulting in Domain 2. (Fig. 3.6A). This bait had a pI of 7.90 and did not autoactivate. The bait strain was mated with the yeast AH109 strain, containing a library of adult *Drosophila melanogaster* cDNA. The mating efficiency was calculated to 1.32 %, which is lower than the protocol recommendations, but the mating reactions still gave positive interactions. Thirteen colonies were found to grow on QDO plates, which indicated that the bait had

successfully been activated by an interaction partner fused to the Gal4 activation domain. PCR reactions were run, using the primers 5´AD LD and 3´AD LD, on all 13 interaction partners. The products were cut with the frequent cutting restriction enzyme *Alu*I (Fig. 3.6B), and this showed that all but one (#6) of the positives seemed to be the same gene. Samples 1, 6, 10, 12 and 13 were sequenced, and BLAST searches with these sequences against the *Drosophila melanogaster* genome were performed. All 5 samples gave the same result; the gene picked up was CG30084, also called *tungus* (*tun*). This gene is cytologically mapped to band 52C4-7 on chromosome arm 2R, and is annotated as alternatively spliced in Fly*Base* (http://flybase.bio.indiana.edu/) with 4 transcripts, PA-PD (Fig. 3.6C); this is not molecularly verified, however. By comparing the sequencing results with the annotated sequence it was shown that it either was transcript PA or PD that interacted with DmG9a. All 5 Y2H-positives also aligned with the same area in TUN, starting approximately 15 aa upstream of the 3 LIM-domains (*Linl-1*, *Isl-1* and *Mec-3*) at the end of the protein, which was identified by using the SMART database on TUN-PD (Fig. 3.6D). All 4 transcripts of *tun* have the same modular design, the difference between them lies in the distance form the first LIM-domain to the cluster of 3 LIM-domains. The LIM domains are tandem zinc-finger structures that function as modular protein-binding interfaces (reviewed in (Kadrmas and Beckerle 2004). This strongly suggests that DmG9a interacts with TUN in the LIM-domains, possibly through the ANK-repeats, which are known to be involved in protein-protein interactions. In addition to the three LIM-domains at the C-terminal end, one PDZ-domain (*PSD-95*, *Dlg* and *ZO-1*), an α-Actinin binding ZASP-like Motif (ZM) and a fourth LIM-domain was identified in the N-terminal end of TUN.

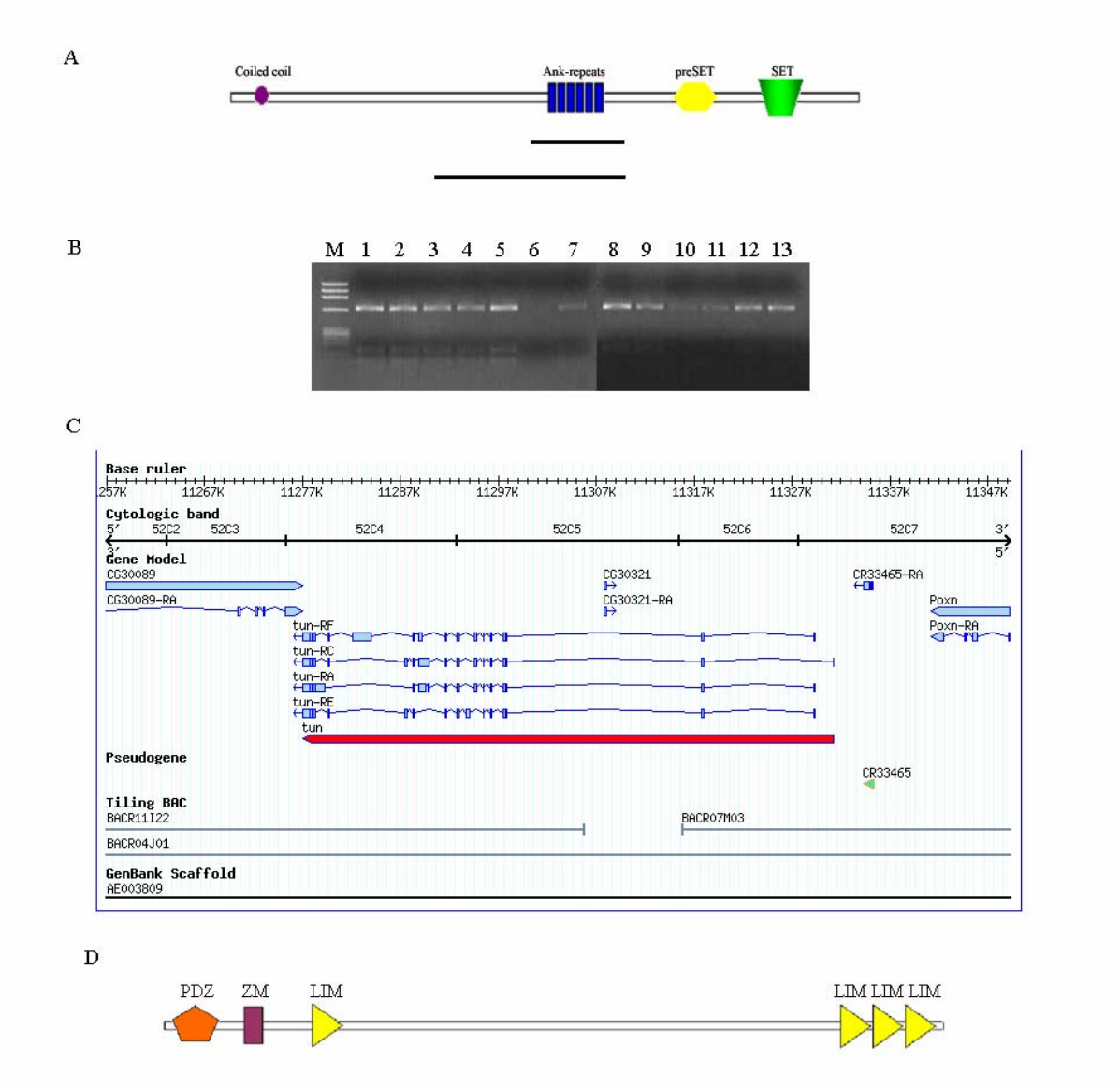

Figure 3.6: Y2h-screen. (A) The two different domains of DmG9a used for the Y2H- screen indicated by lines below the gene. (B) Restriction cutting with *Alu*I of the 13 positives found in the screen. (C) The *Drosophila* gene *tungus* (CG30084), highlighted in red, is located on chromosome arm 2R, cytological band 52C4--7. The gene is annotated with 4 alternative transcripts (adapted from FlyBase (The FlyBase Consortium, 2003)). (D) Overview over the domains of Tungus identified with the SMART database.

On order to verify the interaction further a GST-pulldown experiment was performed. The bait was expressed in bacteria and isolated while the interactant TUN was *in vitro* translated (Fig. 3.7). However, the experimental conditions prevented the assay to be completed, and further adjustments were required to optimize the reactions. This was not done due to lack of time.

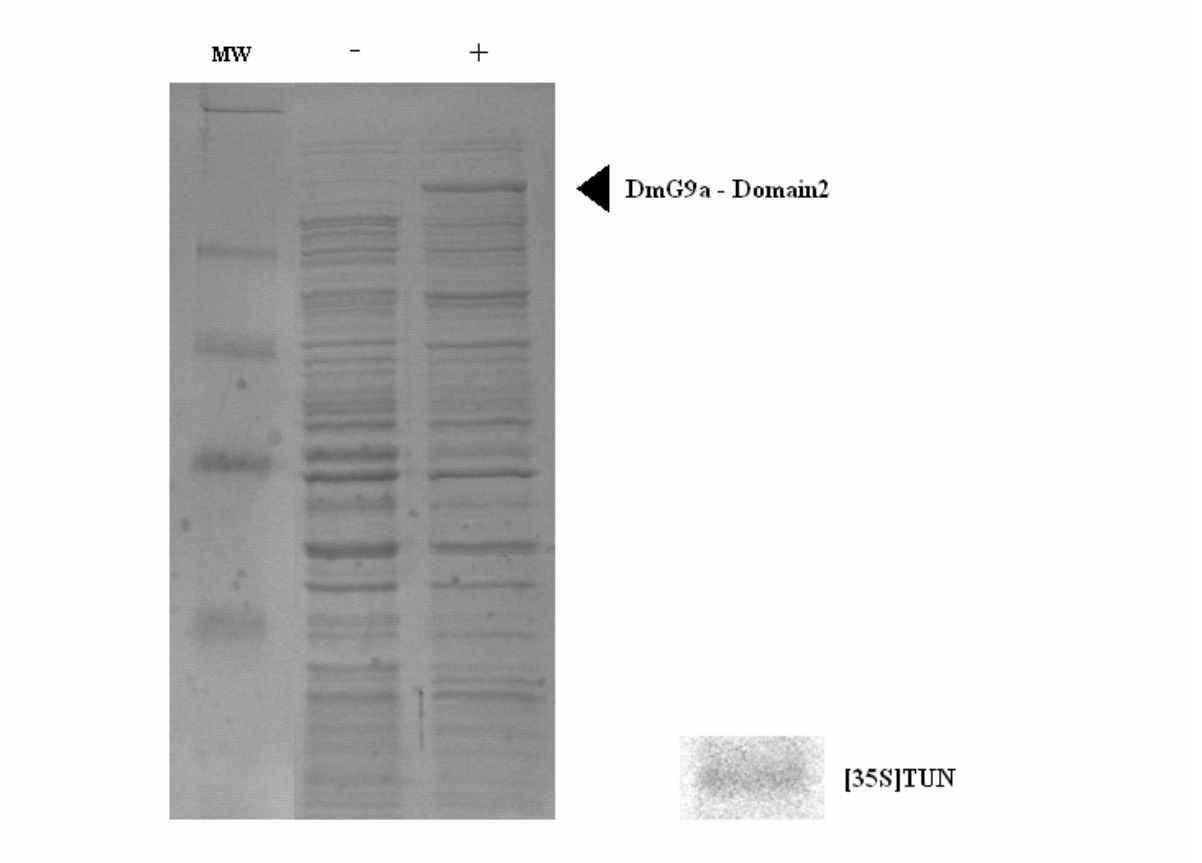

Figure 3.7: Expression in BL21-cells of DmG9a-Domain2 and in vitro translation of TUN.

3.6 Tungus contains an α-Actinin binding domain

The 26 aa long ZM-motif identified in Tungus was named after the human protein ZASP (Z-band alternatively spliced PDZ-motif protein), which is a protein associated with the Z-line in muscles (Faulkner et al. 1999). This is a specialized structure connecting adjacent sarcomeres in muscle cells. ZASP has a homologue in mouse called Cypher; sequence and structure comparison suggests that TUN is a homologue to these two proteins (Figs. 3.8 and 3.7). ZASP/Cypher also has a PDZ-domain in the N-terminal end and three LIM-domains in the C-terminal. PDZ domain is a widely expressed protein-protein interaction domain, with a subgroup of PDZ-LIM proteins (Harris and Lim 2001). PDZ-LIM proteins can be divided in two subclasses; they either contain one or three LIM-domains in the C-terminal end. ZASP and Cypher belongs to the second class, while the proteins ALP (Actinin-associated LIM protein)

and CLP36 (C-terminal LIM domain protein) belongs to the first (Fig. 3.8). All these proteins have been detected in the muscle Z-line in mammals and shown to interact with α-Actinin (Xia et al. 1997; Faulkner et al. 1999; Zhou et al. 1999; Bauer et al. 2000).

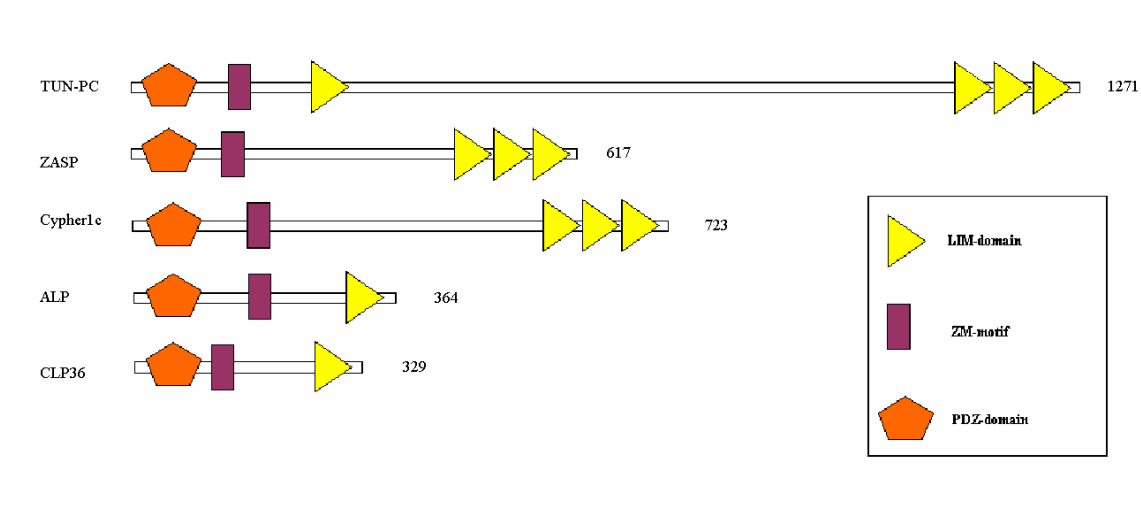

Figure 3.8: PDZ-LIM containing proteins. Modular design of the *Drosophila* protein TUN (splice variant PD, NP_788358), *Homo sapiens* ZASP (CAB46728), ALP (AAB96665), CLP36 (CAC32846) and *Mus musculus* Cypher (splice variant 1c, AAD42950). The proteins are drawn to scale, and the length is indicated to the right. The square shows the different domains identified by the SMART database.

The ZM-motifs and PDZ-domains of the previously identified PDZ-LIM proteins were aligned against TUN, and Figure 3.9 shows that TUN is conserved in both domains. This indicates that is a putative α-Actinin binding protein.

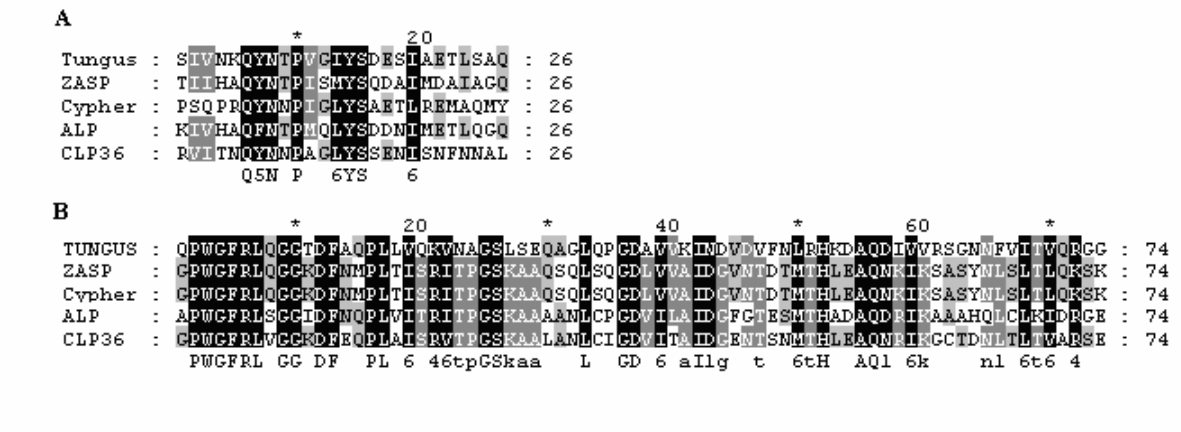

Figure 3.9: Alignment of ZM-motifs and PDZ-domains. Alignment of ZM-motifs (A) and of PDZ-domains (B) of the genes in Figure 6.

3.7 *DmG9a* is abundantly transcribed in ovaries, whereas *tun* is detected in all developmental stages

The expression of *DmG9a* and *tun* was analyzed by reverse transcriptase PCR (RT-PCR). As shown in Fig. 3.10, *DmG9a* transcript was found in all stages of embryos and larvae. 0-3h embryos corresponds to approximately stages 1-7 set by (Campos-Ortega and Hartstein 1985), 3-6 to stages 8-10, 6-9h to stages 11-12 and 9-24h to stages 13-17. In pupae the amount or transcript is very low, while in adult flies with their gonads removed there appears to be no *DmG9a* transcript. The highest amount of transcript is found in ovaries while in testis there is also a moderate amount. *tun* seems to be evenly distributed in all the samples. *Rp49* was used as loading control.

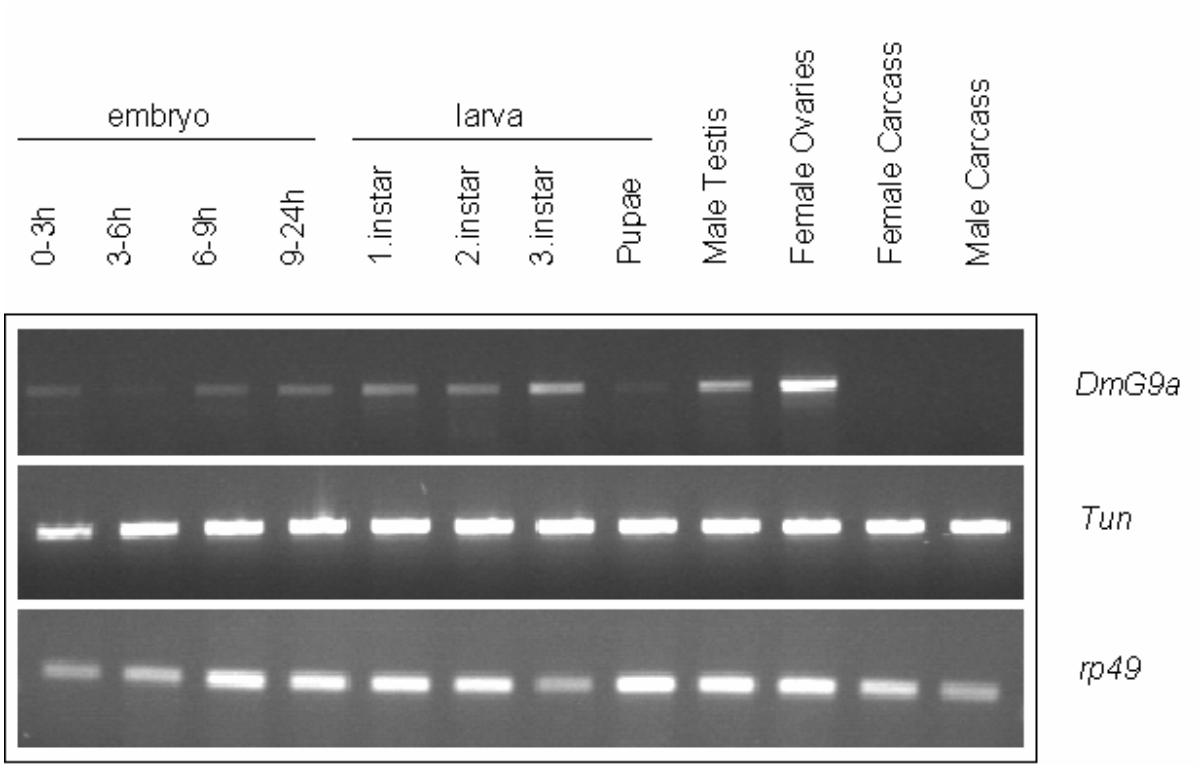

Figure 3.10: RT-PCR expression analysis. Agarose gels stained with ethidium bromide showing cDNA fragments of *DmG9a*, *tun* and *rp49* (positive control) amplified by RT-PCR using gene specific primers. Tissues used are depicted on top. Female and male is adult flies with their sex organs removed.

3.8 *tun* is expressed in *Drosophila* ovaries and embryos

Whole mount *in situ* hybridization towards *tun* transcript showed that the gene is transcribed in the oocyte of stage 9 egg chambers (Fig. 3.11A), at the same time as *DmG9a* transcript is first detected. But while the transcript of *DmG9a* spreads to nurse cells from early stage 10, the *tun* transcript stays in the oocyte from stage 10A through stage 12 (Figs. 3.11B-D). In the following stages, the transcript is no longer detectable (Fig. 3.11E). In stage 5 embryos *tun* transcript is evenly distributed.

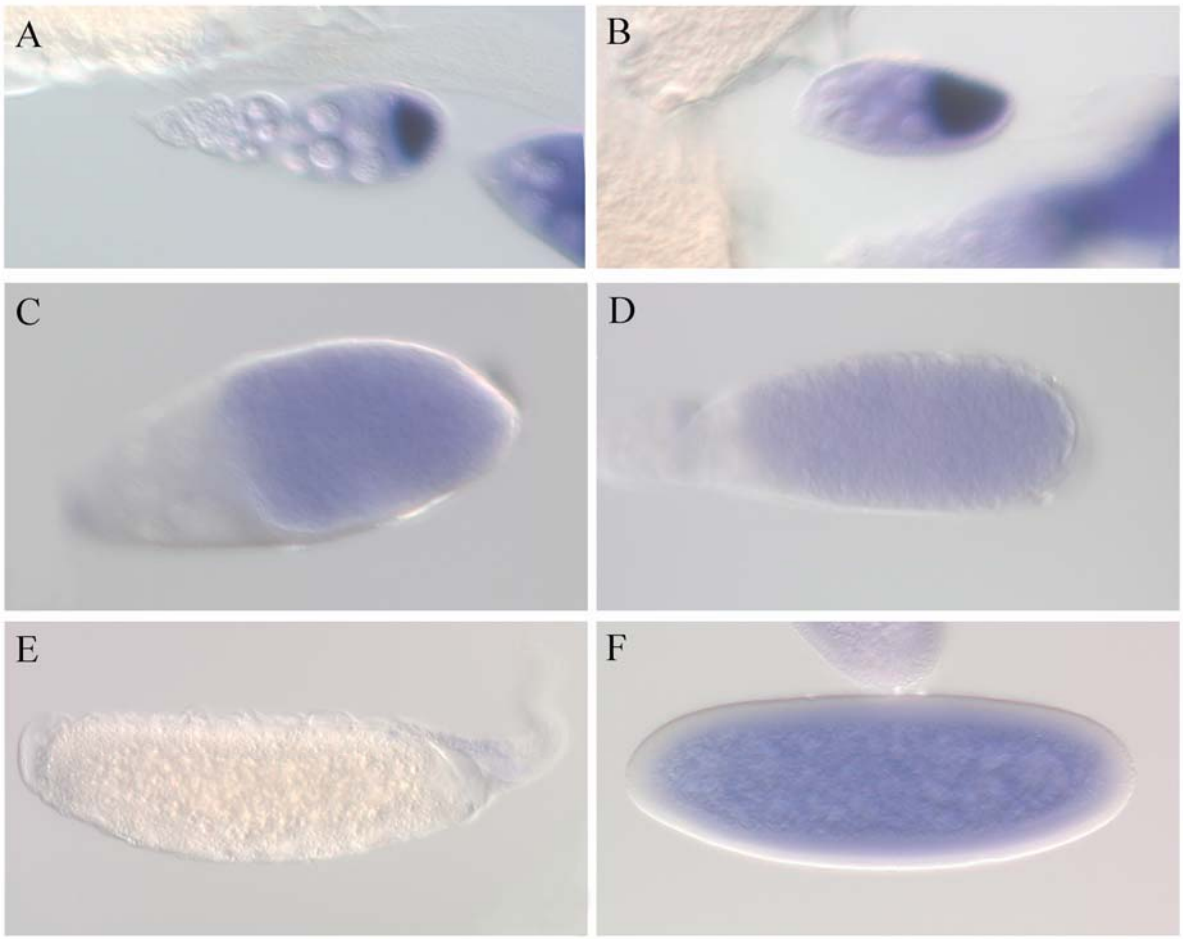

Figure 3.11: Whole mount i*n situ* hybridization of *tun*. (A) Stage 1-9 egg-chambers. (B) Stage 10A egg-chamber. (C) Stage 11 egg-chamber. (D) Stage 12 egg-chamber. (E) Stage 14, mature egg. (F) Stage 5 embryo.

4. Discussion

The evolutionarily conserved SET-domain proteins in eukaryotes have been shown to function as site-specific histone lysine methyl-transferases (HKMTs), and play an important role in regulating chromatin-mediated gene transcriptional activation and silencing. As an approach to gain further insights into the mechanisms of epigenetic gene regulation, a putative histone methyl-transferase was identified and characterized in *Drosophila melanogaster*.

4.1 *DmG9a* belongs to the *Su(var)3-9* family of SET-domain genes

Drosophila melanogaster G9a (*DmG9a*) was identified based on the SET-domain of the well characterized *Drosophila* histone methyl-transferase *Su(var)3-9*, and was named after its homologue in mammals, *G9a,* in this study. *DmG9a*, *Su(var)3-9* and mammalian *G9a* all belongs to the *Su(var)3-9* family of SET-domain genes classified by (Jenuwein et al. 1998). Many of the proteins in this family have been shown to methylate histone H3 lysine (K) 9 (Lachner et al. 2003), including Su(var)3-9 (Czermin et al. 2001; Schotta et al. 2002), suggesting that DmG9a also methylates H3-K9. However, this remains to be shown.

Mammalian G9a is able to add methyl groups to H3-K27 as well as to H3-K9 (Tachibana et al. 2001). The euchromatic H3-K9 specific activity regulated by mouse G9a (MmG9a) has been shown to be essential for early embryogenesis and is involved in the transcriptional repression of developmental genes (Tachibana et al. 2002).

SU(VAR)3-9 is associated with heterochromatin, it interacts with and creates binding sites for HP1 (Jacobs et al. 2001; Schotta et al. 2002), and has been shown to be the main, but not the only H3-K9 HKMT in *Drosophila* (Schotta et al. 2002). In flies, as

well as in other species, it has been observed that some HKMTs have the same target lysine, which raises the issue of redundancy of their functions. This might explain the lack of phenotypic abnormalities in the over-expression lines of *DmG9a*. However, this could also be caused by failure in expressing the vector injected into *Drosophila* lines.

In addition to having the four conserved sequence motifs in its SET-domain, DmG9a also possesses the flanking cysteine rich sequences which have previously been shown to be important for HKMT activity (Rea et al. 2000; Zhang et al. 2002). Thus, *DmG9a* shares all the features of a SET-domain gene, validating its possible function as an HKMT.

4.2 *DmG9a* is a maternal gene required during early embryogenesis

As *MmG9a* had been shown to be embryonic lethal and essential for early development (Tachibana et al. 2002), it was natural to investigate expression pattern of *DmG9a* in *Drosophila* embryos. Whole mount *in situ* hybridization showed that transcript is uniformly present in both stage 4 and 5 embryos. At stage 8, during gastrulation, transcript is seen in the area where mesoderm and ectoderm develops. However, it is difficult to specifically decide which of those cell layers it is transcribed in due to indistinct staining. The presence of *DmG9a* transcript in early embryos was also confirmed in a high-throughput RNA *in situ* protocol used to determine patterns of expression for various genes in *Drosophila* embryonic development (Tomancak et al. 2002). Here, however, staining was only observed from stage 1 through stage 3 embryos. Our RT-PCR results clearly showed, however, that *DmG9a* transcript is present during all stages of embryonic development; this result was also supported by seeing presence of protein at stage 8 (Fig. 3.4E).

The fact that *DmG9a* transcript is present in early embryos, before the maternal to zygotic transition (MZT), shows that the gene is expressed during oogenesis and deposited into the oocyte before fertilization. There is little or no transcription in early *Drosophila* embryos, further supporting the maternal origin of *DmG9a*.

The localization of DmG9a protein in embryos was also investigated. Early embryos showed that DmG9a accumulates in the nuclei in the syncytial blastoderm. During the cellularization process, which starts around stage 5, DmG9a remains associated with the nuclei which are transported towards the perimeter of the embryo, in addition to remain in the uncellularized nuclei in the center. This confirmed that DmG9a is located in the nucleus, which is as expected of an HKMT associated with chromatin. Also the identification of a NLS supports this notion. Later in development, at stage 9, DmG9a becomes associated with ectoderm. This primordial cell layer of the germ band is divided in two clearly defined regions, one that gives rice to the neuroblasts and the ventrolateral epidermis. From the other ectodermal region, the dorsal epidermis and the tracheal placodes develop. The staining of DmG9a shows that it most likely is associated with the first region, also called the neurogenic region.

It has been shown that maintenance of CpG methylation at the Prader-Willi Syndrome Imprinting Center in mouse embryonic stem cells might be dependent on G9a, indicating that the protein has a possible role in imprinting in mice (Xin et al. 2003). As DmG9a also seems to be involved in embryonic development suggests that also DmG9a may be involved in regulation of DNA methylation. In *Drosophila* the overall methylation level of DNA is rather low, with the highest levels found in early embryos (Lyko et al. 2000), which coincides with the expression of *DmG9a*. The function of DNA methylation in *Drosophila*, however, is not well characterized (Field et al. 2004), and its existence is controversial. What possible epigenetic processes DmG9a might be involved in sustains further investigation.

4.3 *DmG9a* is synthesized in nurse cells of *Drosophila* ovaries

Having shown that *DmG9a* is a maternal gene, and that it is expressed in ovaries, it was obvious to investigate the expression pattern in this organ. Transcript is first detected at stage 9, strongly in the oocyte and weakly in nurse cells. In the following stages the amount of transcript seems to be reduced in the oocyte while it increases in nurse cells. One possible interpretation of this observation is that *DmG9a* is deposited into the oocyte mainly at stage 9, and that further increase in oocyte size reduces the overall concentration of transcript here. The increased level of transcript in nurse cells might then reflect a separate function in these cells, indicating that DmG9a is involved in regulation of transcription in both oogenesis and early embryonic development. During the final stage of development is not possible to detect *DmG9a* transcript in the mature egg, in contradiction to what is expected for a maternal gene. One possible explanation is that the probe was unable to penetrate the chorion surrounding the egg. This tough extracellular layer is synthesized during late oogenesis (Margaritis et al. 1980), and is removed from embryos when they are prepared for *in situ* hybridization or immunostaining. However, ovaries are not subject to dechorination as the treatment probably is too harsh and would ruin early egg chambers.

DmG9a protein is detected in egg chambers during early stages of oogenesis, from stage 1 to 8, clearly in follicle cell nuclei and weakly in nurse cell nuclei. However, transcript is not visible in these stages; this could be explained by its low concentration or that the fluorescent antibody is more readily detected than the anti-DIG antibody. Another, less likely possibility is that the protein is synthesized in the germarium, and follows the nuclei throughout oogenesis.

The presence of protein in the nuclei of nurse cells and follicle cells supports the observation from the whole mount *in situ* hybridization assay that DmG9a probably is involved in transcriptional regulation of genes involved in oogenesis, in addition to

the early embryonic development in *Drosophila*. Surprisingly, DmG9a transcript was also found in testis, indicating a role during spermatogenesis. This might reflect a possible role for DmG9a in gonad development or in the development of germline cells.

The transcript of *DmG9a* is first detected in the oocyte at stage 9, but the protein is not detected in the oocyte. Transcriptional products from the oocyte nucleus contribute little to the RNA and protein that will sustain early embryonic development (Mahajan-Miklos and Cooley 1994), so the transcript seen in the oocyte is probably synthesized and transported from the nurse cells.

Around stage 11 DmG9a is shifted from being in the nuclei of nurse cells to accumulate in the cytoplasm. This happens simultaneously as a modified form of apoptosis starts (Buszczak and Cooley 2000), and a rapid transport phase where the remaining cytoplasmic contents of the nurse cells is transferred to the oocyte as a result of a myosin-based contraction. Instead of being transferred to the oocyte, however, DmG9a seems to remain in the nurse cells when they contract. The protein might be degraded at this stage, as the presence of the caspase DREDD is reported from stage 10 (Chen et al. 1998); DREDD is a protein that normally accumulates in cells specified to enter apoptosis.

4.4 DmG9a interacts with the maternal product of *tungus*

As the entire DmG9a protein was considered too big to serve as bait in the Y2H-screen, it was therefore divided into smaller fragments. The ankyrin repeats identified in both DmG9a and MmG9a (Milner and Campbell 1993), are known to function in protein-protein interactions (Bork 1993), and were selected to serve as bait for the Y2H screen. As observed for human G9a and Drosophila Su(var)3-9 (Schotta et al. 2002; Lehner et al. 2004), also this part of DmG9a auto activated in the Y2H-system. This was possible to circumvent by expanding the fragment upstream to give a higher

pI. Only one putative interaction partner, Tungus, was found using this part of DmG9a as bait. The interaction was also confirmed by performing a small scale mating in yeast. There are a couple of explanations for the single hit; the screen was performed using a cDNA library of adult flies, while a more common approach is to screen a library from embryos. Also, only a part of the protein was used in the screen, and other interaction partners might bind to other parts of DmG9a.

tun seems also to be a maternal gene as the transcript is detected in early *Drosophila* embryos. In egg chambers the transcript is visualized in the oocyte from around stage 9, the same stage as *DmG9a*. Here it remains until at least stage 11. *tun* is not detected in the mature egg, probably for the same reason as for *DmG9a*.

4.5 Tungus is a putative α-Actinin binding protein

By analyzing the sequence of Tun, four LIM domains, one PDZ domain and one ZM domain were found. The LIM domain is thought to function as protein interaction modules, and the three in the C-terminal end of TUN probably interacts with the ANK-repeats in DmG9a. Also, PDZ domains function in protein-protein interactions and one subgroup of PDZ proteins have both a PDZ and one or more LIM domains (Harris and Lim 2001), and are regarded as mediators between cytoskeletal structures and signaling cascades. Some of the proteins in this subgroup also have the ZASP motif (ZM), including the homologues to TUN; ZASP, Zypher, ALP and CLP36. All of these proteins have been shown to interact with α-Actinin in mammals (Xia et al. 1997; Faulkner et al. 1999; Zhou et al. 1999; Bauer et al. 2000), and it is therefore reasonable to predict that also TUN can bind the same protein. The single copy version of *α-actinin* in *Drosophila* is alternatively spliced to generate three isoforms that are expressed in larval muscle, adult muscle and non-muscle cells (Roulier et al. 1992), and it is closely related to the vertebrate muscle α-Actinins showing nearly 70 % amino acid identity (Fyrberg et al. 1990). In skeletal and cardiac muscles, α-Actinin is a major component of the Z-disc, where it cross-links actin filaments from

adjacent sarcomeres (Clark et al. 2002). In non-muscle cells it has various roles in the cell, being an adhesion linked, filament-binding scaffold molecule (Otey and Carpen 2004). In *Drosophila* *α*-Actinin is ubiquitously expressed in wild type embryos (Wahlstrom et al. 2004). RT-PCR showed an equal amount of transcript in all tissues investigated, as one would expect from actin, which often is used as a positive control due to its even distribution in all tissues. The possible interaction between TUN and α-Actinin has been verified by the group of Mateos and Baylies (2005), who showed that the two proteins colocalize in the Z-disk, that they share some mutant phenotypes and that TUN show muscle expression in late embryo stages (Mateos and Baylies 2005). As TUN probably binds *α*-Actinin, it is not expected to be present in the nucleus. However, an intron protein GFP trap strategy used to predict nuclear localization of various *Drosophila* proteins (Morin et al. 2001) shows that TUN is expressed in the nuclei of both nurse cells and follicle cells (http://flytrap.med.yale.edu/, Fig. 4.1).

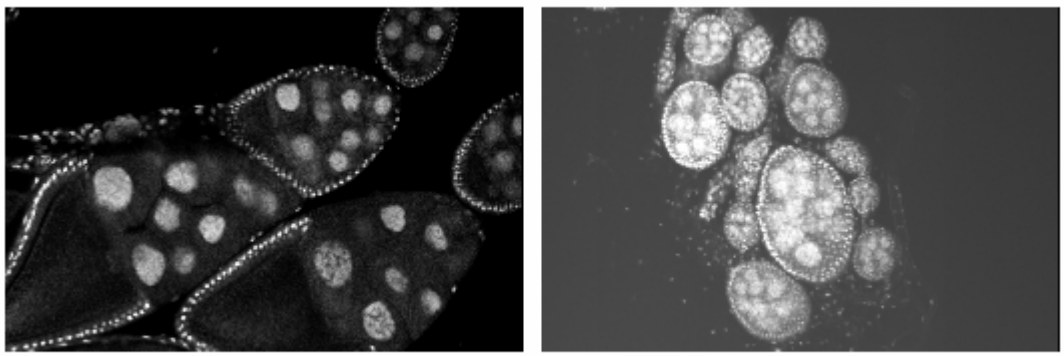

Figure 4.1: TUN expression in ovaries, pictures obtained from FlyTrap (**http://flytrap.med.yale.edu/**)

4.6 TUN might function as a signal for DmG9a

In *Drosophila* ovaries non-muscle α-Actinin is localized in the nurse cell subcortical cytoskeleton, actin cables and ring canals (Wahlstrom et al. 2004), which means that it does not co localize with *tun*. This indicates that TUN is a protein with multiple functions, however, what function it has in ovaries is unknown. The closest characterized homologue to tun is the *C. elegans* protein EAT-1, which has four isoforms (McKeown and Beckerle 2003). One of these isoforms shows nuclear localization, and it is postulated that the protein plays a direct role in signaling (McKeown and Beckerle 2003). Among the multiple roles of non-muscle α-Actinin that has been shown, one is to serve as a scaffold to connect the cytoskeleton to diverse signaling pathways. Taken together, this allows us to speculate that TUN is functioning as a signal that activates DmG9a during *Drosophila* oogenesis, as the colocalization shown for TUN and DmG9a strengthens the notion that they may interact. This is an interesting observation that requires further investigation.

In early stage embryos both *DmG9a, tun* and predominantly non-muscle *α*-Actinin is expressed (Wahlstrom et al. 2004). Cellularization of the *Drosophila* embryo occurs at stage 5, when plasma membrane invaginates to form cleavage furrows between the cortical nuclei, a process known to be associated with actin and myosin II-based contractile apparatus (Warn and Robert-Nicoud 1990; Young et al. 1991). This indicates that there is a connection between DmG9a and TUN also at this stage, and that TUN might function as a signal for DmG9a, which again may regulate genes involved in the cellularization process.

Non-muscle α-Actinin is mainly expressed in young embryos, while muscle α-Actinin is abundant in larvae, pupae and adults (Wahlstrom et al. 2004). TUN show expression both in early and in late embryo stages (Mateos and Baylies 2005), which

means that the protein probably associates with both non-muscle and muscle α-actinin.

To investigate further the interaction between DmG9a and TUN it will be necessary to separate between the different splice variants of *tun*. DmG9a interacted with TUN-PA or Tun-PD, while none of the *in situ* or RT-PCR results for TUN referred to here has been specific for different splice variants. They might be differently expressed, as has been shown for *eat-1* (McKeown and Beckerle 2003).

Another possible function for TUN has also been investigated by Dubnau et al. (2003). This group has shown that a *tun* mutant shows a defective one-day memory, and that it is expressed in mushroom bodies, which are symmetrically paired neuropils in the insect brain that are of critical importance for associative olfactory learning and memory. An interesting feature of DmG9a is that it seems to be expressed in the neurogenic region, the progenitor cells of nervous system. This suggests that DmG9a might function as transcriptional regulator in the early development of the nervous system.

4.7 A possible link between the nucleus and actin

Surprisingly, our results indicate that a putative SET-domain protein possibly interacts with TUN, which is associated with actin. However, another SET-domain protein CG12196 has also showed interaction with TUN (Marianne Stabell, unpublished results). In addition, CG12196 showed interaction with Spire, an actin nucleation factor (Quinlan et al. 2005). Another SET-domain containing protein in mammals, EZH2, has recently been reported as an essential regulator of cellular signaling via ligand-induced actin polymerization in various cell types (Su et al. 2005). Actin, however, has also been proved important in the nucleus as a component of the RNA polymerase II-based transcription machinery (Hofmann et al. 2004). All

these observations support our results, and reveal an interesting link between nuclear transcription factors and actin that needs to be further investigated.

4.8 Future work

Although several observations indicate that DmG9a is a functional HKMT, this has yet to be verified by *in vitro* experiments such as the histone methyltransferase assay described in (Rea et al. 2000). This is, however, an ongoing project in the group, and is essential not only for the demonstration of enzymatic activity, but also in order to elucidate the specificity of the protein. Over the last several years different antibodies has been generated that is specific for different histone residues and methylation states. Knowing what exact residue the DmG9a methylates would give a strong indication for what function it has *in vivo*. Another method for investigation of the function of a gene is through "knock-out" or "knock-down". The first alternative requires creation of a mutant strain, as was tried during this project. An alternative is (Schotta et al. 2002) to perform a "knock-down" experiment using RNAi. In addition to turning off the gene in all cells it would be possible to investigate what happens if the amount of protein is reduced in specific tissues or at specific times. This is especially important in cases where the mutation is recessive lethal of heterozygous lethal.

If a mutant line for *DmG9a* were to be created it would be interesting to investigate the methylation pattern of H3-K9 in early *Drosophila* embryos of *DmG9a; Su(var)3-9* double mutant flies. In single mutant *Su(var)3-9* flies there is a small amount of methylated H3-K9 in embryos (Schotta et al. 2002), and *DmG9a* could be responsible for this methylation.

During this study a promoter analysis assay was started, in order to study in what tissues *DmG9a* is transcribed. Transgenic stocks expressing *lacZ* in the same patterns

as *DmG9a* were obtained, however, due to lack of time the expression pattern was not investigated, and still remains to be done.

In the Y2H-screen TUN was identified as an interaction partner for DmG9a. Although several observations support this relationship, it has not been verified *in vitro*. This is, however, important to do in order to confirm the interaction. GST-pulldown is one method for this purpose that is established in our laboratory; alternatively it is possible to perform a coimmunoprecipitation.

If the interaction was verified a further examination of the expression pattern of *tun* could give supplementary insight into the function of *DmG9a*. Antibodies are under production, and these will help investigate expression pattern for the TUN protein. It would also be possible to study whether DmG9a and TUN colocalize, and in that case; where.

Hopefully the results found from these investigations, and further studies concerning *DmG9a* will contribute to new knowledge and better understanding of epigenetic regulation of transcription.

References

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-95.

Ahmad, K. and S. Henikoff (2002). "Epigenetic consequences of nucleosome dynamics." Cell **111**(3): 281-4.

Ausio, J., D. W. Abbott, X. Wang and S. C. Moore (2001). "Histone variants and histone modifications: a structural perspective." Biochem Cell Biol **79**(6): 693-708.

Bannister, A. J., R. Schneider and T. Kouzarides (2002). "Histone methylation: dynamic or static?" Cell **109**(7): 801-6.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire and T. Kouzarides (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-4.

Bauer, K., M. Kratzer, M. Otte, K. L. de Quintana, J. Hagmann, G. J. Arnold, C. Eckerskorn, F. Lottspeich and W. Siess (2000). "Human CLP36, a PDZ-domain and LIM-domain protein, binds to alpha-actinin-1 and associates with actin filaments and stress fibers in activated platelets and endothelial cells." Blood **96**(13): 4236-45.

Bork, P. (1993). "Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally?" Proteins **17**(4): 363-74.

Brand, A. H. and N. Perrimon (1993). "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes." Development **118**(2): 401-15.

Buszczak, M. and L. Cooley (2000). "Eggs to die for: cell death during Drosophila oogenesis." Cell Death Differ **7**(11): 1071-4.

Campos-Ortega, J. A. and V. Hartstein (1985). *The Embryonic Development of Drosophila melanogaster*. Berlin.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones and Y. Zhang (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**(5595): 1039-43.

Chen, P., A. Rodriguez, R. Erskine, T. Thach and J. M. Abrams (1998). "Dredd, a novel effector of the apoptosis activators reaper, grim, and hid in Drosophila." Dev Biol **201**(2): 202-16.

Cheng, X., R. E. Collins and X. Zhang (2005). "Structural and sequence motifs of protein (histone) methylation enzymes." Annu Rev Biophys Biomol Struct **34**: 267-94.

Cheung, W. L., K. Ajiro, K. Samejima, M. Kloc, P. Cheung, et al. (2003). "Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase." Cell **113**(4): 507-17.

Clark, K. A., A. S. McElhinny, M. C. Beckerle and C. C. Gregorio (2002). "Striated muscle cytoarchitecture: an intricate web of form and function." Annu Rev Cell Dev Biol **18**: 637-706.

Czermin, B., G. Schotta, B. B. Hulsmann, A. Brehm, P. B. Becker, G. Reuter and A. Imhof (2001). "Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila." EMBO Rep **2**(10): 915-9.

Drees, B. L. (1999). "Progress and variations in two-hybrid and three-hybrid technologies." Curr Opin Chem Biol **3**(1): 64-70.

Dubnau, J., A. S. Chiang, L. Grady, J. Barditch, S. Gossweiler, et al. (2003). "The staufen/pumilio pathway is involved in Drosophila long-term memory." Curr Biol **13**(4): 286-96.

Dutnall, R. N. (2003). "Cracking the histone code: one, two, three methyls, you're out!" Mol Cell **12**(1): 3-4.

Engels, W. R. (1983). "The P family of transposable elements in Drosophila." Annu Rev Genet **17**: 315-44.

Faulkner, G., A. Pallavicini, E. Formentin, A. Comelli, C. Ievolella, et al. (1999). "ZASP: a new Z-band alternatively spliced PDZ-motif protein." J Cell Biol **146**(2): 465-75.

Feng, Q., H. Wang, H. H. Ng, H. Erdjument-Bromage, P. Tempst, K. Struhl and Y. Zhang (2002). "Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain." Curr Biol **12**(12): 1052-8.

Field, L. M., F. Lyko, M. Mandrioli and G. Prantera (2004). "DNA methylation in insects." Insect Mol Biol **13**(2): 109-15.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-6.

Fischer, J. A., E. Giniger, T. Maniatis and M. Ptashne (1988). "GAL4 activates transcription in Drosophila." Nature **332**(6167): 853-6.

Fischle, W., Y. Wang and C. D. Allis (2003). "Binary switches and modification cassettes in histone biology and beyond." Nature **425**(6957): 475-9.

Fortini, M. E., M. P. Skupski, M. S. Boguski and I. K. Hariharan (2000). "A survey of human disease gene counterparts in the Drosophila genome." J Cell Biol **150**(2): F23-30.

Frederickson, R. M. (1998). "Macromolecular matchmaking: advances in two-hybrid and related technologies." Curr Opin Biotechnol **9**(1): 90-6.

Friedman, R. and A. L. Hughes (2001). "Pattern and timing of gene duplication in animal genomes." Genome Res **11**(11): 1842-7.

Fyrberg, E., M. Kelly, E. Ball, C. Fyrberg and M. C. Reedy (1990). "Molecular genetics of Drosophila alpha-actinin: mutant alleles disrupt Z disc integrity and muscle insertions." J Cell Biol **110**(6): 1999-2011.

Galau, G. A., D. W. Hughes and L. Dure III (1986). "Abscisic acid induction of cloned cotton late embryogenesis-abundant (*Lea*) mRNAs." Plant Mol Biol **7**(3): 155-170.

Gary, J. D. and S. Clarke (1998). "RNA and protein interactions modulated by protein arginine methylation." Prog Nucleic Acid Res Mol Biol **61**: 65-131.

Giniger, E., S. M. Varnum and M. Ptashne (1985). "Specific DNA binding of GAL4, a positive regulatory protein of yeast." Cell **40**(4): 767-74.

Goll, M. G. and T. H. Bestor (2002). "Histone modification and replacement in chromatin activation." Genes Dev **16**(14): 1739-42.

Harris, B. Z. and W. A. Lim (2001). "Mechanism and role of PDZ domains in signaling complex assembly." J Cell Sci **114**(Pt 18): 3219-31.

Hofmann, W. A., L. Stojiljkovic, B. Fuchsova, G. M. Vargas, E. Mavrommatis, et al. (2004). "Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II." Nat Cell Biol **6**(11): 1094-101.

Horn, P. J. and C. L. Peterson (2002). "Molecular biology. Chromatin higher order folding--wrapping up transcription." Science **297**(5588): 1824-7.

Hudson, A. M. and L. Cooley (2002). "Understanding the function of actin-binding proteins through genetic analysis of Drosophila oogenesis." Annu Rev Genet **36**: 455-88.

Isenberg, I. (1979). "Histones." Annu Rev Biochem **48**: 159-91.

Jacobs, S. A., S. D. Taverna, Y. Zhang, S. D. Briggs, J. Li, J. C. Eissenberg, C. D. Allis and S. Khorasanizadeh (2001). "Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3." Embo J **20**(18): 5232-41.

Jenuwein, T. (2001). "Re-SET-ting heterochromatin by histone methyltransferases." Trends Cell Biol **11**(6): 266-73.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.

Jenuwein, T., G. Laible, R. Dorn and G. Reuter (1998). "SET domain proteins modulate chromatin domains in eu- and heterochromatin." Cell Mol Life Sci **54**(1): 80-93.

Jiang, J., T. Hoey and M. Levine (1991). "Autoregulation of a segmentation gene in Drosophila: combinatorial interaction of the even-skipped homeo box protein with a distal enhancer element." Genes Dev **5**(2): 265-77.

Kadrmas, J. L. and M. C. Beckerle (2004). "The LIM domain: from the cytoskeleton to the nucleus." Nat Rev Mol Cell Biol **5**(11): 920-31.

Keegan, L., G. Gill and M. Ptashne (1986). "Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein." Science **231**(4739): 699-704.

Khorasanizadeh, S. (2004). "The nucleosome: from genomic organization to genomic regulation." Cell **116**(2): 259-72.

Kouzarides, T. (2002). "Histone methylation in transcriptional control." Curr Opin Genet Dev **12**(2): 198-209.

Kubicek, S. and T. Jenuwein (2004). "A crack in histone lysine methylation." Cell **119**(7): 903-6.

Lachner, M. and T. Jenuwein (2002). "The many faces of histone lysine methylation." Curr Opin Cell Biol **14**(3): 286-98.

Lachner, M., D. O'Carroll, S. Rea, K. Mechtler and T. Jenuwein (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature **410**(6824): 116-20.

Lachner, M., R. J. O'Sullivan and T. Jenuwein (2003). "An epigenetic road map for histone lysine methylation." J Cell Sci **116**(Pt 11): 2117-24.

Lawrence, P. A. (1992). The Making of a Fly: The Genetics of Animal Design, Blackwell Scientific Publications.

Lee, D. Y., C. Teyssier, B. D. Strahl and M. R. Stallcup (2005). "Role of protein methylation in regulation of transcription." Endocr Rev **26**(2): 147-70.

Lehner, B., J. I. Semple, S. E. Brown, D. Counsell, R. D. Campbell and C. M. Sanderson (2004). "Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region." Genomics **83**(1): 153-67.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.

Lyko, F., B. H. Ramsahoye and R. Jaenisch (2000). "DNA methylation in Drosophila melanogaster." Nature **408**(6812): 538-40.

Mahajan-Miklos, S. and L. Cooley (1994). "Intercellular cytoplasm transport during Drosophila oogenesis." Dev Biol **165**(2): 336-51.

Margaritis, L. H., F. C. Kafatos and W. H. Petri (1980). "The eggshell of Drosophila melanogaster. I. Fine structure of the layers and regions of the wild-type eggshell." J Cell Sci **43**: 1-35.

Margueron, R., P. Trojer and D. Reinberg (2005). "The key to development: interpreting the histone code?" Curr Opin Genet Dev **15**(2): 163-76.

Mateos, J. and M. K. Baylies (2005). "Role of the PDZ-LIM protein Tungus in muscle fusion and assembly." A. Dros. Res. Conf. 46 "Program and Abstracts. 46th Annual Drosophila Research Conference, San Diego, CA, 2005" **46**(209B): 157.

McBride, A. E. and P. A. Silver (2001). "State of the arg: protein methylation at arginine comes of age." Cell **106**(1): 5-8.

McGinnis, W. and R. Krumlauf (1992). "Homeobox genes and axial patterning." Cell **68**(2): 283-302.

McKeown, C. R. and M. C. Beckerle (2003). "The EAT-1 ALP/Enigma proteins are required for muscle cell adhesion in C. elegans." International Worm Meeting 2003 **114**.

Milner, C. M. and R. D. Campbell (1993). "The G9a gene in the human major histocompatibility complex encodes a novel protein containing ankyrin-like repeats." Biochem J **290 (Pt 3)**: 811-8.

Min, J., Q. Feng, Z. Li, Y. Zhang and R. M. Xu (2003). "Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase." Cell **112**(5): 711-23.

Morgan, T. H., A. H. Sturtevant, H. J. Muller and C. B. Bridges (1915). The Mechanism of Mendelian Heredity. New York, The Maple Press York PA.

Morin, X., R. Daneman, M. Zavortink and W. Chia (2001). "A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila." Proc Natl Acad Sci U S A **98**(26): 15050-5.

Muller, H. J. (1930). "Types of visible variations induced by X-rays in *Drosophila*." J Genet **22**: 299-334.

Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." Nature **287**(5785): 795-801.

Orlando, V. and R. Paro (1995). "Chromatin multiprotein complexes involved in the maintenance of transcription patterns." Curr Opin Genet Dev **5**(2): 174-9.

Otey, C. A. and O. Carpen (2004). "Alpha-actinin revisited: a fresh look at an old player." Cell Motil Cytoskeleton **58**(2): 104-11.

Paro, R. and D. S. Hogness (1991). "The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila." Proc Natl Acad Sci U S A **88**(1): 263-7.

Peterson, C. L. and M. A. Laniel (2004). "Histones and histone modifications." Curr Biol **14**(14): R546-51.

Pirrotta, V. (1998). "Polycombing the genome: PcG, trxG, and chromatin silencing." Cell **93**(3): 333-6.

Quinlan, M. E., J. E. Heuser, E. Kerkhoff and R. D. Mullins (2005). "Drosophila Spire is an actin nucleation factor." Nature **433**(7024): 382-8.

Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature **406**(6796): 593-9.

Robinson, D. N. and L. Cooley (1997). "Examination of the function of two kelch proteins generated by stop codon suppression." Development **124**(7): 1405-17.

Roulier, E. M., C. Fyrberg and E. Fyrberg (1992). "Perturbations of Drosophila alpha-actinin cause muscle paralysis, weakness, and atrophy but do not confer obvious nonmuscle phenotypes." J Cell Biol **116**(4): 911-22.

Rubin, G. M. and E. B. Lewis (2000). "A brief history of Drosophila's contributions to genome research." Science **287**(5461): 2216-8.

Sambrook, J. and D. W. Russell (2001). Molecular cloning: a laboratory manual. New York, Cold Spring Harbor Laboratory Press.

Schotta, G., A. Ebert, V. Krauss, A. Fischer, J. Hoffmann, S. Rea, T. Jenuwein, R. Dorn and G. Reuter (2002). "Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing." Embo J **21**(5): 1121-31.

Schultz, J., F. Milpetz, P. Bork and C. P. Ponting (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-64.

Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole and R. A. Casero (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." Cell **119**(7): 941-53.

Simon, J. A. and J. W. Tamkun (2002). "Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes." Curr Opin Genet Dev **12**(2): 210-8.

Sims, R. J., 3rd, K. Nishioka and D. Reinberg (2003). "Histone lysine methylation: a signature for chromatin function." Trends Genet **19**(11): 629-39.

Spradling, A. C. (1993). Developmental genetics of oogenesis. The Development of *Drosophila melanogaster*. M. Bate and A. M. Arias. New York, Cold Spring Harbor Laboratory Press. **I:** 1-70.

Spradling, A. C., D. Stern, A. Beaton, E. J. Rhem, T. Laverty, N. Mozden, S. Misra and G. M. Rubin (1999). "The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes." Genetics **153**(1): 135-77.

Spradling, A. C., D. M. Stern, I. Kiss, J. Roote, T. Laverty and G. M. Rubin (1995). "Gene disruptions using P transposable elements: an integral component of the Drosophila genome project." Proc Natl Acad Sci U S A **92**(24): 10824-30.

St Johnston, D. (2002). "The art and design of genetic screens: Drosophila melanogaster." Nat Rev Genet **3**(3): 176-88.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-5.

Su, I. H., M. W. Dobenecker, E. Dickinson, M. Oser, A. Basavaraj, et al. (2005). "Polycomb group protein ezh2 controls actin polymerization and cell signaling." Cell **121**(3): 425-36.

Tachibana, M., K. Sugimoto, T. Fukushima and Y. Shinkai (2001). "Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3." J Biol Chem **276**(27): 25309-17.

Tachibana, M., K. Sugimoto, M. Nozaki, J. Ueda, T. Ohta, et al. (2002). "G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis." Genes Dev **16**(14): 1779-91.

Tautz, D. and C. Pfeifle (1989). "A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback." Chromosoma **98**(2): 81-5.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.

Tomancak, P., A. Beaton, R. Weiszmann, E. Kwan, S. Shu, et al. (2002). "Systematic determination of patterns of gene expression during Drosophila embryogenesis." Genome Biol **3**(12): RESEARCH0088.

Turner, B. M. (2002). "Cellular memory and the histone code." Cell **111**(3): 285-91.

Vermaak, D., K. Ahmad and S. Henikoff (2003). "Maintenance of chromatin states: an open-and-shut case." Curr Opin Cell Biol **15**(3): 266-74.

Wahlstrom, G., V. P. Lahti, J. Pispa, C. Roos and T. I. Heino (2004). "Drosophila non-muscle alpha-actinin is localized in nurse cell actin bundles and ring canals, but is not required for fertility." Mech Dev **121**(11): 1377-91.

Warn, R. M. and M. Robert-Nicoud (1990). "F-actin organization during the cellularization of the Drosophila embryo as revealed with a confocal laser scanning microscope." J Cell Sci **96 (Pt 1)**: 35-42.

Weigmann, K., R. Klapper, T. Strasser, C. Rickert, G. Technau, H. Jackle, W. Janning and C. Klambt (2003). "FlyMove--a new way to look at development of Drosophila." Trends Genet **19**(6): 310-1.

Weisberg, R. A., L. W. Enquist, C. Foeller and A. Landy (1983). "Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda." J Mol Biol **170**(2): 319-42.

Xia, H., S. T. Winokur, W. L. Kuo, M. R. Altherr and D. S. Bredt (1997). "Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs." J Cell Biol **139**(2): 507-15.

Xin, Z., M. Tachibana, M. Guggiari, E. Heard, Y. Shinkai and J. Wagstaff (2003). "Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center." J Biol Chem **278**(17): 14996-5000.

Young, P. E., T. C. Pesacreta and D. P. Kiehart (1991). "Dynamic changes in the distribution of cytoplasmic myosin during Drosophila embryogenesis." Development **111**(1): 1-14.

Zhang, X., H. Tamaru, S. I. Khan, J. R. Horton, L. J. Keefe, E. U. Selker and X. Cheng (2002). "Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase." Cell **111**(1): 117-27.

Zhou, Q., P. Ruiz-Lozano, M. E. Martone and J. Chen (1999). "Cypher, a striated muscle-restricted PDZ and LIM domain-containing protein, binds to alpha-actinin-2 and protein kinase C." J Biol Chem **274**(28): 19807-13.

Aagaard, L., G. Laible, P. Selenko, M. Schmid, R. Dorn, et al. (1999). "Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31." Embo J **18**(7): 1923-38.

Aagaard, L., M. Schmid, P. Warburton and T. Jenuwein (2000). "Mitotic phosphorylation of SUV39H1, a novel component of active centromeres, coincides with transient accumulation at mammalian centromeres." J Cell Sci **113 (Pt 5)**: 817-29.

Appendix 1

Experiment	Comment	Primer name	Sequence	Tm (oC)
Control of insert *DmG9a*		2995leftRNAi	5´- TGTCGCACTTCTCGTTCATC – 3´	50.9
Control of insert *DmG9a*		2995seq3R	5´- ACCATCATTGTGCAGCGTAG – 3´	51.0
In situ probe/RT-PCR *tun*		IS30084left	5´ - GAAAGGGCCGTAAGTCAGTT - 3´	50.2
In situ probe/RT-PCR *tun*		IS30084right	5´ - ATGGAAGTGTTTGCCAATGG – 3´	52.3
Loading control RT-PCR		rp49L	5´- TGACCATCCGCCCAGCATAC - 3´	57.9
Loading control RT-PCR		rp49R	5´- TCTCGCCGCAGTAAACGC - 3´	54.8
Overexpression		2995cDNAattB1	5´- GGGGACAAGTTTGTACAAAAAAGCAGGCTGCAAGATGACGGACTTTGTT - 3´	79.0
"		cg2995Y2H4attB2	5´ - GGGGACCACTTTGTACAAGAAAGCTGGGTACGCGTGTCCAATTTTCTCCGT – 3´	81.8
Probe Southern		2995intron/biotin	5´- AAGCACAAGGACAAGGAGGA - 3´	51.2
"		2995intron	5´- CGCATTTTGTATGTGGTTCG - 3´	51.6
Promoter analysis		2995prom/EcoRI	5´ – GGCGAATTCTTTTCTTACATTCTACGTGG – 3´	42.3
"		2995prom/BamHI	5´ – CAAGGATCCCTTGCCTTAATTGTCACTAT - 3´	41.4
RT-PCR *DmG9a*		2995right(RNAi)	5´ – ACTATCCAGGCAGGAGCAGA - 3´	51.3
"		2995left(RNAi)	5´ – GATGAACGAGAAGTGCGACA - 3´	50.9
Screening of Y2H positives and *In vitro* translation of tun for GST-pulldown		5´AD LD-Insert Screening Amplimer	5´- CTATTCGATGATGAAGATACCCCACCAAACCC – 3´	67.0
"		5´AD LD-Insert Screening Amplimer	5´- GTGAACTTGCGGGGTTTTTCAGTATCTACGATT - 3´	66.1
Y2H-bait	Domain2	cg2995Y2H2attB1	5´- GGGGACAAGTTTGTACAAAAAAGCAGGCTACGAACTGCGCTGTGAGTTCGAA – 3´	81.4
"	Domain1 and Domain2	cg2995Y2H3attB2	5´- GGGGACCACTTTGTACAAGAAAGCTGGGTAAACCACAAAGGTACGGAGGCC – 3´	80.9
"	Domain1	cg2995Y2H3attB1	5´- GGGGACAAGTTTGTACAAAAAAGCAGGCTCATATCCCGCGGGGTAAACTCTG – 3´	81.3